

ALEXANDRE CHAVES FARIA

Dissertação apresentada ao Programa de Pós-graduação em Geografia - Tratamento da Informação Espacial da Pontifícia Universidade Católica de Minas Gerais, como requisito parcial à obtenção do grau de Mestre.

Área de concentração: Análise espacial.

Orientador: Prof. Dr. Cláudio Caetano Machado

Co-orientador: Prof. Dr. Leônidas Conceição Barroso

Mestrando: Alexandre Chaves Faria

Determinação das Coordenadas Geográficas de Imóveis Urbanos: Desenvolvimento de Aplicativo Computacional - Abordagem exploratória de *Geocoding* em GIS

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE
MINAS GERAIS

**Belo Horizonte
2005**

FICHA CATALOGRÁFICA

F224d	<p>Faria, Alexandre Chaves Determinação das coordenadas geográficas de imóveis urbanos: desenvolvimento de aplicativo computacional: abordagem exploratória de <i>geocoding</i> em GIS / Alexandre Chaves Faria. Belo Horizonte, 2005. 144f.</p> <p>Orientador: Cláudio Caetano Machado Co-orientador: Leônidas Conceição Barroso Dissertação (Mestrado) – Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Tratamento da Informação Espacial. Bibliografia</p> <p>1. Sistemas de informação geográfica. 2. Coordenadas geográficas. 3. Imóveis – Banco de dados. I. Machado, Cláudio Caetano. II. Barroso, Leônidas Conceição. III. Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Tratamento da Informação Espacial. IV. Título.</p> <p style="text-align: right;">CDU: 91:681.3</p>
-------	--

Título: **Determinação das Coordenadas Geográficas de Imóveis Urbanos:**
Desenvolvimento de Aplicativo Computacional - Abordagem Exploratória de
Geocoding em GIS.

Autor: Alexandre Chaves Faria

Aprovada em:

Examinadores:

Prof. Dr. Cláudio Caetano Machado (Orientador) – PUC Minas

Prof. Dr. Leônidas Conceição Barroso (Co-orientador) – PUC Minas

Prof. Dr. João Francisco de Abreu – PUC Minas

Prof. Dr. Clodoveu Augusto Davis Júnior – PUC Minas

AGRADECIMENTOS

À Deus.

Agradeço à Deus pela provisão diária, pelo cuidado permanente e pela capacitação concedida em todos os momentos de minha vida e, em especial, na conquista deste objetivo.

À família.

À Ione, Juliana, Yvonne, Luiza e Ricardo, esposa e filhos queridos que no decorrer do tempo relevaram todos os momentos de convívio subtraídos em prol da realização deste projeto.

Aos professores.

Dirijo meus agradecimentos à todos os professores do curso de Mestrado em Tratamento da Informação Espacial, em especial ao meu orientador professor doutor Cláudio Caetano Machado e ao meu co-orientador professor doutor Leônidas Conceição Barroso. Foram todos eles exemplos de dedicação, excelência e amor à causa do saber.

À CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

Pelo importante incentivo concedido através da liberação da bolsa de estudos.

À COPASA, colegas e amigos.

À Companhia de Saneamento de Minas Gerais - COPASA MG, pela oportunidade de realização deste projeto e aos meus colegas e amigos Regina Darck Cançado e Márcio Antônio Jorge pelo incentivo e apoio constantes.

“Inexiste no mundo coisa mais bem distribuída que o bom senso, visto que cada indivíduo acredita ser tão bem provido dele que mesmo os mais difíceis de satisfazer em qualquer outro aspecto não costumam desejar possuí-lo mais do que já possuem.”

(DESCARTES – O Discurso do Método)

RESUMO

A crescente demanda por informações oferecidas pelo GIS (*Geographic Information System*) gerou uma grande busca por dados espaciais em formato digital como insumo para a produção das análises e resultados desejados. Essa dissertação visa o desenvolvimento de um aplicativo computacional para a produção de uma base de dados espaciais em formato digital dos clientes da COPASA – Companhia de Saneamento de Minas Gerais – no contexto da cidade de Belo Horizonte. Para tal, foram determinadas as coordenadas geográficas dos endereços dos clientes da companhia, aliando-se a funcionalidade de geocodificação do GIS com a metodologia de localização de endereços da COPASA. Essa abordagem teve por objetivo obter um melhor resultado do processo de geocodificação no que se refere ao percentual total de endereços geocodificados (*address-matching rate*) e à solução de alguns problemas comuns observados na geocodificação de endereços. Além do cadastro de clientes da COPASA, foi utilizado como base de dados de referência o arquivo digital de eixos de logradouros da prefeitura municipal de Belo Horizonte. O desenvolvimento do aplicativo GIS foi estruturado em duas fases: a fase de especificação e a fase de projeto físico. Na fase de especificação foi adotada a metodologia de análise essencial de sistemas. O ambiente utilizado para a implementação do aplicativo foi o ARC/INFO 8.3 e a linguagem de programação AML – *Arc Macro Language*, específica do ambiente ARC/INFO. Em que pese o custo de digitalização e preparo da base de dados de referência para representar a estrutura de rotas da metodologia de localização de imóveis da COPASA, os resultados obtidos foram plenamente satisfatórios. Uma alta taxa de geocodificação e a solução a alguns problemas comuns da geocodificação de endereços foram alcançadas pela abordagem de geocodificação empregada.

ABSTRACT

The growing demand for information offered by GIS (Geographic Information System) has generated a great need for digital-format spatial data as resources for the production of analysis and results wanted. This dissertation has sought a computational application development in order to create a digital and spatial database of the clients of COPASA – Minas Gerais Sanitation Company - in the context of Belo Horizonte city. The main points to get this were determining the geographical coordinates of the company clients' addresses, and allying the geocoding GIS functionality to the property localization methodology of COPASA. This approach had the objective of getting a better result on the geocoding process concerning the address-matching rate and the solution of common problems noticed on address geocoding. Besides the client's archives, a digital centerline database, provided by the city hall, was used as a database of reference. The development of GIS application was structured in two phases: the specification phase and the physical project phase. In the specification phase it was adopted the essential system analysis methodology. ARC/INFO 8.3 was the environment chosen for the implementation of the application and AML – Arc Macro Language – the programming language, specific for ARC/INFO environment. In spite of the digitalization and the reference database preparation costs, the results obtained in representing the route structures of property localization methodology of COPASA were fully satisfactory. A high address-matching rate and the solution for some usual address geocoding problems were obtained by the employed geocoding approach.

LISTA DAS FIGURAS

Figura 2.1.1 – Modelo holístico e funcional de um sistema de informações geográficas	28
Figura 2.1.2 – Camadas espaciais do GIS	32
Figura 2.1.3 – Hierarquia de construção de feições complexas.	33
Figura 2.1.4 – Propriedades das feições dos objetos geográficos	34
Figura 2.1.5 – Estrutura do grid no formato raster.	36
Figura 2.1.6 – Estrutura do modelo vetorial arco-nó.	38
Figura 2.1.7 – Atributos associados a pontos georreferenciados.	38
Figura 2.1.8 – Atributos associados a arcos georreferenciados.	39
Figura 2.1.9 – Atributos associados a polígonos georreferenciados.	39
Figura 2.1.10 – Topologia básica do arco	41
Figura 2.2.1 - Estrutura básica da faixa de endereços nos eixos de logradouros dos arquivos TIGER.	49
Figura 2.2.2 – Casos de cólera mapeados pelo Dr. Snow em Londres em 1854.	52
Figura 2.2.3 – Processo de geocodificação de endereços em uma base de referência de trechos de logradouros.	53
Figura 2.2.4 – Padrão de disposição da numeração ao longo do logradouro determinado pelo sentido do trecho.	55
Figura 2.2.5 - Exemplo do processo de geocodificação de endereços em uma base de referência de trechos de logradouros.	55
Figura 2.2.6 – Fator de afastamento na geocodificação de endereços.	56
Figura 2.2.7 – Fator de encurtamento na geocodificação de endereços.	57
Figura 3.1.1 – Exemplo de eixo de logradouro com seus atributos associados.	75
Figura 3.1.2 – Divisão em setores das rotas de leitura de consumo e entrega de faturas do distrito de serviços sudoeste da COPASA em Belo Horizonte.	85
Figura 3.1.3 – Exemplo de uma rota de leitura de consumo e entrega de faturas da COPASA em Belo Horizonte.	86
Figura 3.1.4 – Exemplo de localização de um imóvel pela metodologia de leitura de consumo e entrega de faturas da COPASA.	87
Figura 3.2.1 - Modelagem do problema na base de dados de referência.	89
Figura 3.2.2 – Croqui da rota de leitura 39 do setor 15.	91
Figura 3.2.3 – Exemplo do processo de cadastramento do setor-rota-face nos trechos de logradouros (setor 15 rota 39).	92
Figura 3.2.4 – Exemplo de trecho de logradouro com sentido incorreto.	93
Figura 3.2.5 – Ilustração do modelo de <i>geocoding</i> adotado que trabalha com subconjuntos dos arquivos de endereços alvo e de referência.	95
Figura 3.3.1 – Mapa da rota 54 setor 15.	103
Figura 3.3.2 – Mapa dos clientes geocodificados da rota 01 setor 15.	104
Figura 3.3.3 – Relatório de consistência do SIURB contendo as faces cadastradas em duplicidade.	105
Figura 3.3.4 – Relatório de consistência do SIURB contendo os trechos de logradouro com sentido invertido.	

.....	105
Figura 3.4.1 – Estatística do processo de geocoding do ArcGis 8.3 para o setor 15 dos clientes da COPASA.	107
Figura 3.4.2 – Exemplo de uma região cujos trechos de referência não correspondem ao croqui da rota (rota 1549).	3.4.2 – Observação do Posicionamento Dentro dos Limites de Quadras 109
3.4.2 – Observação do Posicionamento Dentro dos Limites de Quadras 110
Figura 3.4.3 – Erro de posicionamento devido a um erro construtivo da rota. 110
Figura 3.4.4 – Erro de posicionamento devido ao ângulo agudo acentuado da esquina. 111
Figura 3.4.5 – Erro de posicionamento devido ao fator de afastamento inadequado. 112
Figura 3.4.6 – Resultado da geocodificação da face 150364 contendo numeração irregular. 113

LISTA DAS TABELAS

Tabela 1.3.1 – População, área total e densidade demográfica de Minas Gerais e Belo Horizonte.	20
Tabela 1.3.2 – População e área dos bairros de contexto do projeto de dissertação.	20
Tabela 2.1.1 – Fontes comuns de erros encontrados na utilização do GIS.	45
Tabela 3.1.1 – Atributos básicos dos trechos de logradouros.....	73
Tabela 3.1.2 - Níveis que compõem a base de dados do SIGRAE.	77
Tabela 3.2.1 – Fatores de afastamento utilizados.	98
Tabela 3.2.2 – Fatores de encurtamento utilizados.	98

SUMÁRIO

CAPÍTULO I - INTRODUÇÃO.....	12
1.1 – INTRODUÇÃO.....	12
1.2 - JUSTIFICATIVAS.....	15
1.3 - OBJETIVOS.....	18
1.4 - CONTEXTO.....	19
1.5 - A COPASA.....	23
CAPÍTULO II – MÉTODOS E TÉCNICAS.....	26
2.1 – GEOGRAPHIC INFORMATION SYSTEMS - GIS.....	26
2.1.1 – <i>Introdução</i>	26
2.1.2 – <i>Componentes de um GIS</i>	29
2.1.3 – <i>Abrangência do GIS</i>	30
2.1.4 – <i>Camadas espaciais</i>	32
2.1.5 – <i>Objetos Espaciais</i>	33
2.1.6 – <i>Propriedades das feições</i>	33
2.1.7 – <i>Modelos de implementação de dados do GIS</i>	36
2.1.8 – <i>Topologia em GIS</i>	40
2.1.9 – <i>Qualidade dos dados espaciais</i>	41
2.2 – GEOCODIFICAÇÃO.....	46
2.2.1 – <i>Introdução</i>	46
2.2.2 – <i>Geocodificação de endereços</i>	50
2.2.3 – <i>O processo de geocodificação de endereços</i>	53
2.2.4 – <i>Os erros na geocodificação de endereços</i>	58
2.2.5 – <i>Tratamento dos registros não geocodificados</i>	61
2.2.6 – <i>Ferramentas para aumentar a taxa de correspondência</i>	62
2.3 – METODOLOGIA E TÉCNICAS UTILIZADAS.....	65
2.3.1 – <i>Especificação</i>	65
2.3.2 – <i>Projeto físico</i>	67
CAPÍTULO III – DESENVOLVIMENTO DO TRABALHO.....	69
3.1 - INTRODUÇÃO.....	69
3.1.1 – <i>Descrição da realidade</i>	69
3.1.2 – <i>O Problema da geocodificação de endereços</i>	79
3.1.3 – <i>A metodologia de leitura de consumo e entrega de faturas da COPASA</i>	82
3.2 – PROCEDIMENTOS EXECUTADOS.....	88
3.2.1 – <i>Fase de Especificação</i>	88
3.2.2 – <i>Fase de Projeto Físico</i>	100
3.3 – PRODUTOS FINAIS.....	102
3.3.1 – <i>Mapas de rota</i>	102

3.3.2 - Mapas dos clientes georreferenciados.....	104
3.3.3 - Relatórios de consistência	105
3.4 – ANÁLISE E VALIDAÇÃO DOS RESULTADOS	106
3.4.1 – Observação da Taxa de Geocodificação.....	106
3.4.2 – Observação do Posicionamento Dentro dos Limites de Quadras.....	110
3.4.3 – Observação do Tratamento do Problema da Numeração Irregular.....	113
CAPÍTULO IV – CONCLUSÃO	114
REFERÊNCIAS BIBLIOGRÁFICAS.....	118
ANEXOS.....	120
ANEXO 1 – MODELO AMBIENTAL DO SIURB	121
ANEXO 2 – MODELO COMPORTAMENTAL	124
ANEXO 3 – DESCRIÇÃO DAS BASES DE DADOS	126
ANEXO 4 - DIAGRAMA FÍSICO DO SISTEMA.....	133
ANEXO 5 - CÓDIGO AML DO PROGRAMA DE GEOCODIFICAÇÃO DOS CLIENTES	138
ANEXO 6 – NÚMEROS E INDICADORES DA COPASA	143

Capítulo I - Introdução

1.1 – Introdução

Todo acontecimento ocorre no espaço e no tempo. Portanto, nossa percepção do mundo é inerentemente espacial e temporal: objetos possuem uma localização, e os eventos são entrelaçados em uma linha do tempo. Como forma de expressar nossas percepções para que elas possam ser comunicadas a outros, usamos modelos. Construimos modelos não só para comunicar, mas para nós mesmos compreendermos o mundo. Ao manipularmos os modelos descobrimos como eles funcionam e como o mundo que eles representam pode parecer no futuro. Podemos, então, sistematicamente experimentar com os modelos para descobrir meios de evitar um futuro indesejável ou alcançar um futuro desejável. ¹

Todo fato, acontecimento ou fenômeno geográfico possui uma dimensão temporal e espacial no campo da realidade que necessita ser tratada para a produção do conhecimento geográfico. Já no campo do conhecimento, a escolha do melhor tratamento, metodologia e representação desse fato depende, em grande parte, de suas características e, em outra parte, também das perguntas que se deseja responder ou do conhecimento que se quer alcançar.

Na geografia existem algumas questões básicas nas quais devem estar fundamentadas todo o estudo e pesquisa do fato geográfico: “*onde é?*”, “*o que é?*”, “*quando foi?*”, “*o que mudou?*”, “*por onde ir?*”, “*qual o padrão?*”, “*por que ocorre?*” e “*o que acontece se...?*” são algumas perguntas que devem invariavelmente ser respondidas pelo investigador. ²

No centro de todas essas questões, ressalta-se a importância da localização do fato geográfico: “*onde é?*”. Quer seja para a simples descrição do fenômeno, ou para o estudo mais aprofundado de suas causas, efeitos, correlações e previsões, é primordial que o fenômeno a ser estudado esteja associado a um sistema de localização, tal como suas coordenadas geográficas.

¹ FOTHERINGHAM, S. & WEGENER, M., "Spatial models and GIS: new potential and new models". Ed. Taylor & Francis Group, 2000, p3-6.

² Notas de aula da disciplina “Métodos de análise espacial” do Prof. Dr. João Francisco de Abreu - Programa de Pós-graduação em Geografia – Tratamento da Informação Espacial, PUC Minas, 2003.

Cabe, portanto, ao geógrafo valer-se de um senso crítico quanto aos fenômenos localizados, fazendo correlações e buscando explicá-los. MEYNIER (1971) ³ assinala que “... os primeiros traços da inteligência básicos para quem pretende trabalhar um dia com a geografia são o senso de orientação e uma aguda capacidade de observação”. De fato, o senso de orientação e a capacidade de observação são imprescindíveis ao geógrafo, mas a representação da realidade através de modelos e o uso de tratamentos da informação mais adequados ampliam a percepção desta mesma realidade e ajudam a encontrar as melhores soluções e respostas ao problema geográfico investigado.

Na geografia, o tratamento descritivo, literário ou mesmo verbal da informação geográfica tem sido, desde as suas origens, muito utilizado para caracterizar os fatos geográficos e representá-los de maneira a se alcançar um maior e melhor conhecimento do mundo real. Desde as primeiras concepções geográficas gregas, atribuídas a Homero, nas narrativas da *Iliada* e da *Odisséia*, provavelmente no século VIII a.C., passando por geógrafos como Humboldt e Ritter no início do século XIX, até os dias atuais, o tratamento descritivo dos fatos geográficos tem sido largamente empregado. Todavia, outras formas de representação foram e devem ser perseguidas para a ampliação da capacidade do homem de entender a realidade.

Uma dessas formas foi, sem dúvida, a construção de mapas. As representações geográficas estão entre as mais antigas, tendo suas raízes nas necessidades das sociedades mais primitivas. Conforme LONGLEY (2001) ⁴, “os mapas devem ter sua origem nos rabiscos que os primitivos fizeram sobre a terra ou nas paredes das cavernas, muito antes da linguagem se tornar suficientemente sofisticada para comunicar as informações equivalentes presentes no discurso”.

A ciência tem buscado, neste sentido, em todo o campo do conhecimento, as melhores e mais adequadas formas de representação da realidade para a tomada de decisões e para as ações que alicerçam a evolução da sociedade humana.

³ AMORIM FILHO, O. B., “A Geografia e as qualidades necessárias ao geógrafo”, PUF, 1971. Traduzido de MEYNIER, A: “Guide de l’Étudiant en Géographie, Paris.

⁴ LONGLEY, PAUL A. et al. “Geographic Information Systems and Science”, Ed. John Wiley & Sons, Ltd.,2001.

O avanço da tecnologia, em geral, e da tecnologia da informação, em particular, possibilita implementar novas formas de representação da realidade, que vão ao encontro da ampliação da capacidade de entendimento do homem sobre esta mesma realidade. Assim, o desenvolvimento da ciência da computação, em um primeiro momento, e da ciência da informação espacial, mais recentemente, fornece exemplos de novas alternativas para a escolha do melhor tratamento, metodologia e representação da informação geográfica por parte do investigador.

Procuramos apresentar uma discussão e a aplicação de uma metodologia de representação da localização de endereços urbanos, através de um tratamento computacional da informação e sua representação digital por meio de pares de coordenadas geográficas em um sistema de informações geográficas. Certamente essa forma de representação da realidade não esgota todas as possibilidades de entendimento dos fatos, eventos e objetos que se apresentam na natureza e que ocorrem na superfície da Terra, mas busca disponibilizar ao investigador científico ferramentas poderosas para a organização da massa de dados existente, normalmente grande, com vistas à obtenção de uma tipologia ou um padrão, além de auxiliá-lo na simulação e na predição.

1.2 - Justificativas

O conhecimento de onde se encontram os cidadãos em um determinado município é uma informação valiosa tanto do ponto de vista do estabelecimento de estratégias e prioridades no atendimento de suas necessidades, como do controle das ações por parte do poder público. Também no universo das empresas privadas, tem sido grande a demanda por informações que possibilitem o conhecimento de onde se encontram seus clientes, tanto os reais como os potenciais. O estabelecimento de estratégias corretas de marketing, a descoberta de novos mercados, a maximização de investimentos direcionados a mercados específicos, são objetivos buscados tanto em empresas públicas como privadas.

Existe, de fato, um número cada vez maior de aplicações sócio-econômicas que requerem por parte do pesquisador a manipulação de base de dados contendo o endereço como o atributo que caracteriza espacialmente um determinado evento. O avanço das tecnologias de GIS e os computadores têm dado grande impulso na tomada de decisões de planejamento urbano e políticas governamentais, entre outras, onde o endereço individual está se tornando um padrão para a investigação espacial. Os recursos já existem, mas é essencial a produção de uma base de dados que indique a localização dos cidadãos no espaço geográfico, associada ao maior número possível de informações descritivas pertinentes, resguardadas as questões éticas e de privacidade decorrentes.

Para utilizar toda a natureza espacial do endereço é comumente necessário representá-lo como um ponto de coordenadas x e y associado ao registro que contém o endereço. Devido a esta necessidade, quase a totalidade das ferramentas de GIS possui a funcionalidade de geocodificar (*address geocoding*) – processo de associação de um endereço descritivo qualquer a um ponto em um mapa – e de realizar consultas espaciais com base nos resultados.

Neste sentido, foi escolhida a Companhia de Saneamento de Minas Gerais – COPASA – como objeto de estudo, uma vez que a mesma possui um rico cadastro de seus clientes, tanto reais como potenciais, bem como uma base cartográfica planimétrica digital do município de Belo Horizonte. Os clientes, entretanto, não possuem suas coordenadas geográficas determinadas, o que inviabiliza diversas análises espaciais essenciais ao planejamento de ações, tanto no âmbito operacional como no âmbito mercadológico.

No âmbito operacional, a existência de uma base de dados que contenha a distribuição espacial de seus clientes possibilita, por exemplo, pelo uso da propriedade topológica de continência, o conhecimento do consumo de água em uma determinada área geográfica. Este conhecimento é valioso para se medir e controlar a perda de água no processo de distribuição, localizar ligações clandestinas, planejar melhorias e expansões na rede de distribuição, entre outros.

No âmbito mercadológico, vários estudos de mercado podem ser levados avante pelo conhecimento das características sócio-econômicas dos clientes, indicando áreas de segmentação de mercado, ocorrência de clientes potenciais, localização de postos de atendimento, entre outros.

Para que estes estudos e análises tenham qualidade, é necessário que a base de dados geográfica dos clientes seja a mais confiável e exata possível. Para tanto, o processo de geocodificação de endereços assume um papel importante na produção de uma base de dados geográfica com qualidade, visto que os resultados podem ser aceitáveis ou não, dependendo de como ele é executado.

A funcionalidade de geocodificação utilizada pelas ferramentas de GIS baseia-se, frequentemente, na correspondência dos endereços entre dois bancos de dados, onde normalmente os atributos de relacionamento são: o tipo do logradouro, o nome do logradouro e o número do imóvel. Este método habitual apresenta um percentual total de endereços geocodificados bastante variado, dependendo da qualidade das bases de dados envolvidas. Na maioria dos casos encontram-se correspondências perfeitas entre 25% a 75% dos registros do arquivo de endereços (Drummond, 1995).⁵ Muito raramente se atinge percentuais mais elevados, principalmente em conjuntos de dados muito amplos, devido às diferenças nos padrões de endereços, erros de digitação das bases de dados, desatualização da base de dados de referência, entre outros. Também alguns endereços são geocodificados erroneamente, ou mesmo não geocodificados, quando a realidade de campo não segue o padrão estabelecido pelo método, como é o caso da existência de logradouros com uma numeração irregular.

⁵ DRUMMOND, William J., "Address matching", *Journal of the American Planning Association*; Spring95, Vol. 61 Issue 2, p240,11p. Este estudo se refere à realidade americana, onde, em geral, as irregularidades como a ocorrência de logradouros homônimos e com numeração irregular são raras.

Uma vez que não existe um padrão mínimo para a taxa de correspondência e os valores aceitáveis podem variar, dependendo do estudo que se deseja fazer (Harries, 1999), ⁶ este trabalho de pesquisa buscou desenvolver um sistema de informações geográficas no qual foi implementada uma adaptação do processo habitual de geocodificação, com o intuito de minimizar os problemas normalmente encontrados na geocodificação de endereços e alcançar uma melhor taxa de correspondência para a produção de uma base de clientes geocodificados a mais confiável e exata possível.

⁶ HARRIES, K., 1999, "Mapping Crime: Principles and Practice", (Washington: US Department of Justice).

1.3 - Objetivos

A dissertação visa o desenvolvimento de um aplicativo computacional, doravante chamado Sistema de Informações Urbanas (SIURB), que realizará a geocodificação dos endereços dos clientes da COPASA, ou seja, a partir de um banco de dados alfanumérico contendo os endereços dos clientes e de uma base digital urbana georreferenciada dos trechos de logradouros, irá obter as coordenadas geográficas planas em projeção UTM (Universal Transverse Mercator) dos endereços e a representação por pontos destes endereços em um arquivo digital.

Além disso, este trabalho propõe uma abordagem exploratória para a utilização da funcionalidade de *geocoding* - presente em vários pacotes de GIS - baseada na metodologia de leitura de consumo e entrega de faturas da Companhia de Saneamento de Minas Gerais – COPASA.

1.4 - Contexto

O contexto de implantação do sistema SIURB abrange, inicialmente, todo o município de Belo Horizonte (figura 1.3.1), uma vez que o mesmo possui o cadastro de eixos de logradouros digitalizado no ambiente GIS. Sua implantação, entretanto, poderá ser estendida a qualquer localidade que possua a base de dados digital dos eixos de logradouros disponível. A metodologia também pode ser aplicada a todos os municípios operados pela COPASA, bem como a qualquer outro município do Brasil que disponha de uma base de dados digital dos eixos de logradouros e de uma metodologia semelhante à adotada pela COPASA nas rotinas de leitura de consumo e entrega de faturas.

Para efeito da análise e validação dos resultados, este trabalho faz uso do cadastro de clientes (banco de dados comercial) da COPASA em Belo Horizonte, compreendendo os bairros: Alto dos Caiçaras, Aparecida, Aparecida - Sétima Seção, Bom Jesus, Caiçara Adelaide, Caiçaras, Dom Cabral, Ermelinda, Jardim Montanhês, Minas Brasil, Monsenhor Messias, Nova Cachoeirinha, Nova Esperança, Padre Eustáquio, Pedro II e Santo André (figura 1.3.2). As tabelas 1.3.1 e 1.3.2 mostram dados de área e população para as regiões de contexto deste trabalho.

Este contexto foi definido em função da COPASA possuir uma base urbana georreferenciada atualizada dos eixos de logradouros e por ser uma região bem representativa da realidade das localidades atendidas pela COPASA. Neste sentido, esta área apresenta a estrutura de aglomerados subnormais, arruamentos com numeração irregular, bem como abrange todo um setor de planejamento de serviços de leitura de consumo e entrega de faturas da COPASA.⁷

No anexo 6 são mostrados alguns números e indicadores da COPASA para uma melhor compreensão do contexto de atuação da empresa.

⁷ Um setor de planejamento de serviços de leitura de consumo e entrega de faturas é um conjunto de rotas de percurso urbano para visitaç o de im veis por parte de leituristas para se efetuar a leitura de consumo de  gua e/ou a entrega de faturas.

Região	População	Área (km²)	Densidade demográfica
Minas Gerais	17.891.494	586.931,64	30,50
Belo Horizonte	2.238.526	331,57	6.764,96

Tabela 1.3.1 – População, área total e densidade demográfica de Minas Gerais e Belo Horizonte.⁸

Bairros do contexto de validação	População	Área (Ha)
Alto dos Caiçaras	1.330	12,557
Aparecida	6.934	56,240
Aparecida - Sétima Seção	6.017	63,114
Bom Jesus	4.058	30,067
Caiçara Adelaide	10.892	78,210
Caiçaras	14.299	260,490
Dom Cabral	4.813	94,656
Ermelinda	2.350	23,725
Jardim Montanhês	3.351	154,835
Minas Brasil	1.802	21,082
Monsenhor Messias	4.819	40,986
Nova Cachoeirinha	8.978	64,057
Nova Esperança	3.824	36,698
Padre Eustáquio	33.579	311,630
Pedro II	5.460	61,482
Santo André	10.553	87,107
Total	123.059	1396,936

Tabela 1.3.2 – População e área dos bairros de contexto do projeto de dissertação.⁹⁸ Fonte: IBGE, Censo demográfico 2000.⁹ Fonte: Prodabel, Censo demográfico 2000.

Região de contexto do projeto de dissertação para implantação do SIURB, Belo Horizonte - MG

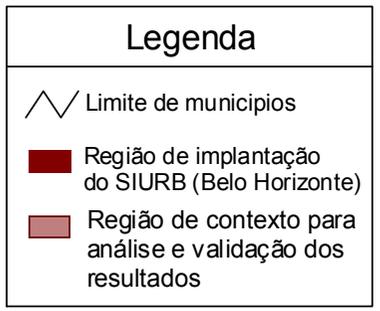
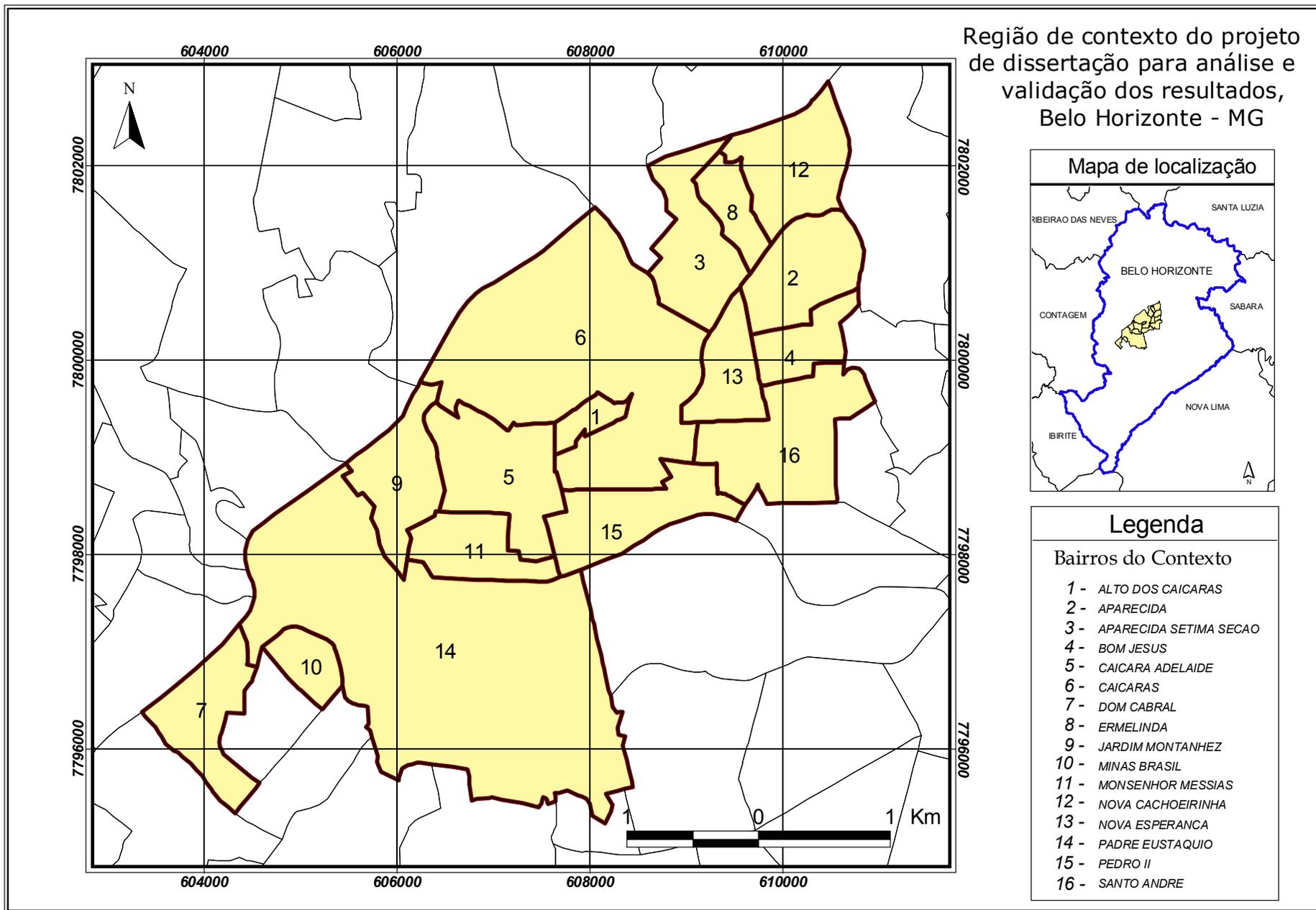


Figura 1.3.1 – Mapa de contexto do projeto de dissertação para a implantação do sistema SIURB.

Fonte: PUC Minas - Laboratório de geodemografia (IBGE 2000)

Figura 1.3.2 – Mapa de contexto do projeto de dissertação para a validação dos resultados.



Fonte: PUC-Minas - Laboratório de geodemografia (IBGE 2000)

1.5 - A COPASA ¹⁰

A COPASA tem suas origens na criação da Companhia Mineira de Águas e Esgotos (COMAG) no ano de 1963. Fundada para planejar, executar e explorar diretamente os serviços de água potável e coleta de esgotos sanitários no âmbito do Estado de Minas Gerais, a COMAG surgiu como resposta à necessidade de novas formas administrativas e financeiras capazes de dar maior flexibilidade operacional aos órgãos do setor de saneamento e reverter o quadro precário das condições de saneamento básico das cidades mineiras. Em 1960 somente 19% da população mineira era servida por rede de água e 23% por rede de coleta de esgoto.¹¹

Embora criada para impulsionar o setor de saneamento em todo o Estado, a COMAG se ressentia da falta de recursos do Tesouro Estadual mantendo-se por algum tempo restrita à administração de uns poucos sistemas de produção e distribuição de água, até que na década de 70 a empresa recebe da Superintendência do Desenvolvimento do Nordeste (SUDENE) a responsabilidade de operar e administrar diversas cidades do norte de Minas, passando a contar com recursos financeiros do Banco Nacional de Habitação (BNH), para implantar sistemas de água em cidades localizadas em outras regiões, como Santos Dumont, Araxá, Betim e Patos de Minas.

Em janeiro de 1973, a Companhia Mineira de Água e Esgotos (COMAG) assina um convênio com o BNH, o governo mineiro e o Banco de Crédito Real de Minas Gerais, por intermédio do qual fixa as condições para colocar em prática o Programa de Abastecimento de Água do Estado nos moldes preconizados pelo Plano Nacional de Saneamento (PLANASA).

A empresa acelerou sua consolidação quando esta, em maio de 1973, encampou o Departamento Municipal de Água e Esgoto (DEMAE) do município de Belo Horizonte, beneficiando-se significativamente pela transferência de uma importante infra-estrutura em termos de equipamentos e recursos humanos.

Somente em 14 de novembro de 1974 a COMAG passa a chamar-se Companhia de Saneamento de Minas Gerais (COPASA). A partir daí, inicia-se para a COPASA uma fase de

¹⁰ Fonte: ROSA, E. S. & equipe, Saneamento Básico em Belo Horizonte: Trajetória em 100 anos, Rona Editora, FJP/CEHC, Belo Horizonte, 1996.

¹¹ Fonte: FIBGE – SEPLAN/MG

expansão acelerada no sentido de alcançar as metas definidas pelo PLANASA até 1980: atender com água potável a mais de 80% da população urbana de pelo menos 80% das cidades e prestar serviços de coleta de esgoto a no mínimo 50% da população urbana da Região Metropolitana de Belo Horizonte e das principais cidades mineiras. Durante a década de 70 predomina na COPASA uma postura expansionista, mas ainda revelando uma situação frágil do ponto de vista do equilíbrio econômico-financeiro.

Em 1979, numa experiência inédita, a COPASA passa a ser mutuária direta do Banco Mundial. Para absorver os investimentos advindos da negociação direta com o Banco Mundial, a COPASA precisou investir em sua modernização, aprimorando o seu sistema comercial, financeiro, de desenvolvimento, de recursos humanos e de planejamento operacional.

A história do saneamento no Brasil mostra que muitas administrações municipais conferiam ao departamento de água e esgoto um caráter provinciano e casuístico, em que o favorecimento político era a moeda comum. À medida que ampliava seus serviços, a COPASA começou a ser vista como empresa de utilidade pública onde o modelo personalista não podia mais ser sustentado. O recadastramento de clientes foi um instrumento essencial utilizado em 1980 para uma mudança de filosofia comercial, uma mudança de abordagem política da comunidade e de instauração de uma regra de mercado entre empresa e sociedade. O cadastro de clientes passa de uma posição secundária para uma posição de destaque e crítica para a sobrevivência da empresa. Por três meses a COPASA mobilizou-se para a revisão de seu cadastro, atraindo com isso a atenção da mídia e provocando acirrados debates.

Os anos 80 são marcados pela instabilidade econômica e financeira, de um lado, mas também pela consolidação da filosofia de trabalho baseada na preservação ambiental e na qualidade de água, filosofia esta que persiste até os dias atuais.

Um novo desafio surge na virada dos anos 80 para os 90, quando surge a possibilidade de recuperação do poder político municipal pela nova Constituição brasileira. O novo quadro afeta particularmente as companhias de saneamento do País, dando início a uma concorrência cada vez mais acirrada pela obtenção das concessões de operação de sistemas de abastecimento de água e esgoto sanitário e mesmo para a manutenção das já existentes. Inaugura-se, então, uma nova fase no relacionamento entre o poder concedente e as

concessionárias de serviço público no País.

Na COPASA, uma amostra da nova forma de conduzir a política de saneamento básico se deu com a celebração de um convênio com a prefeitura de Belo Horizonte em julho de 1994, estabelecendo uma parceria e co-responsabilidade no desenvolvimento dos serviços de água e esgotos. Definiu-se como prioridade a ação em vilas e favelas, de forma a incorporá-las à cidade formal. A renovação da concessão, ocorrida em 2000, representou a consolidação deste modelo de parceria na gestão dos serviços de abastecimento de água e tratamento de esgoto sanitário.

Também uma fase de parceria técnica inicia-se com o convênio assinado em 1991 entre a COPASA e a Prefeitura de Belo Horizonte, para o intercâmbio de base de dados georreferenciadas, onde a Prefeitura se compromete a ceder sua base de dados referente ao mapeamento urbano e a COPASA sua base de dados referente às redes de água e esgoto sanitário. Em 2004 este convênio é novamente celebrado com a ampliação do número de órgãos públicos participantes e o estabelecimento de ações coordenadas de compatibilização de endereços entre os diversos órgãos.¹²

Ao longo de sua existência, a COPASA vivenciou e superou muitos desafios e obstáculos que a fizeram adaptar-se às novas situações e realidades apresentadas. Atualmente, entretanto, não basta só adaptar-se. É necessário fazê-lo velozmente. A realidade de nossos dias exige decisões certas e rápidas, que nem sempre são possíveis em face de uma realidade complexa e repleta de conexões. Neste sentido, a COPASA hoje busca o auxílio de ferramentas que auxiliem não só na ampliação de seu poder de reação, mas principalmente na sua capacidade de previsão e antecipação dos fatos. Foi pensando assim que, em 1995, implantou um projeto de desenvolvimento de um sistema de informações geográficas para a gestão de suas redes de distribuição de água e coletoras de esgoto, com vistas a utilizar o potencial de análise, simulação e previsão características de um GIS. Também agora, busca utilizar esses benefícios na espacialização de seus clientes.

¹² A compatibilização de endereços visa uma padronização dos endereços entre as empresas de modo a se obter uma identificação unívoca dos endereços de seus clientes reais e potenciais, além da natural consistência e enriquecimento de seus cadastros.

Capítulo II – Métodos e Técnicas

2.1 – Geographic Information Systems - GIS

2.1.1 – Introdução

O termo *Geographic Information Systems* – GIS pode gerar alguma confusão. Em um extremo, ele é muito utilizado para descrever um determinado programa de computador ou pacote de software, como o ArcView ou MapInfo.¹³ Tais programas são projetados para, de forma genérica, armazenar, consultar, integrar, recuperar, mostrar e modelar dados espaciais. Eles se caracterizam por apresentar uma linguagem de comandos ou um sistema de janelas que realizam diversas funções que requerem certo grau de conhecimento por parte dos operadores que deles se utilizam. Cabe aos usuários decidir quais as operações apropriadas para a manipulação e análise dos dados que devem ser realizadas para a solução de um determinado problema. Em um outro extremo, GIS pode significar um sistema de informações especificamente projetado para realizar um determinado subconjunto de operações selecionadas para auxiliar em algumas atividades, tais como a produção automática de mapas, o gerenciamento do tráfego urbano, entre outras. Nestes sistemas, temos algumas operações e uma estrutura de dados muito bem definidas e específicas para a aplicação projetada.

Estes dois extremos, na verdade, nos dão somente uma pálida idéia da complexidade e abrangência do GIS. Para um melhor entendimento, transcrevemos algumas definições encontradas em algumas publicações sobre o assunto:

“Um sistema de informações geográficas é uma facilidade para a preparação, apresentação, e interpretação de fatos que pertencem à superfície da Terra. Esta é uma definição ampla... uma definição consideravelmente mais estreita, contudo, é mais freqüentemente empregada. Em uma linguagem mais comum, um sistema de informações geográficas ou GIS é uma configuração de hardware e software de computador, especificamente projetada para a aquisição, manutenção e uso de dados cartográficos”.¹⁴

¹³ ArcView e MapInfo são marcas registradas da ESRI e Mapinfo Corporation, respectivamente.

¹⁴ TOMLINS, D. “Geographic Information Systems and Cartographic Modeling”, Englewood Cliffs, NJ: Prentice Hall, 1990, pag xi.

“Um sistema de informações geográficas (GIS) é um sistema de informações que é projetado para trabalhar com dados referenciados por coordenadas espaciais ou geográficas. Em outras palavras, um GIS é tanto um banco de dados com capacidades específicas para dados espacialmente referenciados, como um conjunto de operações para manipulação de dados... em um sentido, um GIS pode ser pensado como um mapa de mais alta ordem”.¹⁵

“Um GIS é uma coleção organizada de hardware, software, dados geográficos, e pessoas projetadas para eficientemente capturar, armazenar, manipular, analisar, e apresentar todas as formas de informações geograficamente referenciadas”.¹⁶

TOMLINSON (2003)¹⁷ assinala que o GIS é uma tecnologia particularmente horizontal no sentido de que ele tem uma larga faixa de aplicações na indústria e no campo intelectual. Por esta razão, há uma tendência de se resistir a definições simplistas. Ele propõe assim, um modelo para que se possa alcançar uma compreensão comum a respeito do que está envolvido quando falamos de GIS. A figura 2.1.1 mostra um modelo holístico funcional de um sistema de informações geográficas. Este modelo nos mostra que o GIS armazena dados espaciais com atributos logicamente associados em um banco de dados espacial, onde funções analíticas são controladas interativamente por um operador humano para gerar os produtos de informação desejados. Este modelo será utilizado para se examinar mais detalhadamente cada componente.

¹⁵ STAR, J. & ESTES, J., “Geographic Information Systems: An Introduction, Englewood Cliffs, NJ: Prentice Hall, 1990, pag. 2

¹⁶ ESRI, “Understanding GIS: The Arc/INFO Method”, Redlands, CA: ESRI, 1990, pag. 1.2

¹⁷ TOMLINSON, Roger F., “Thinking About GIS: geographic information system planning for managers” Redlands, CA: ESRI, 2003, pag. 1-5.

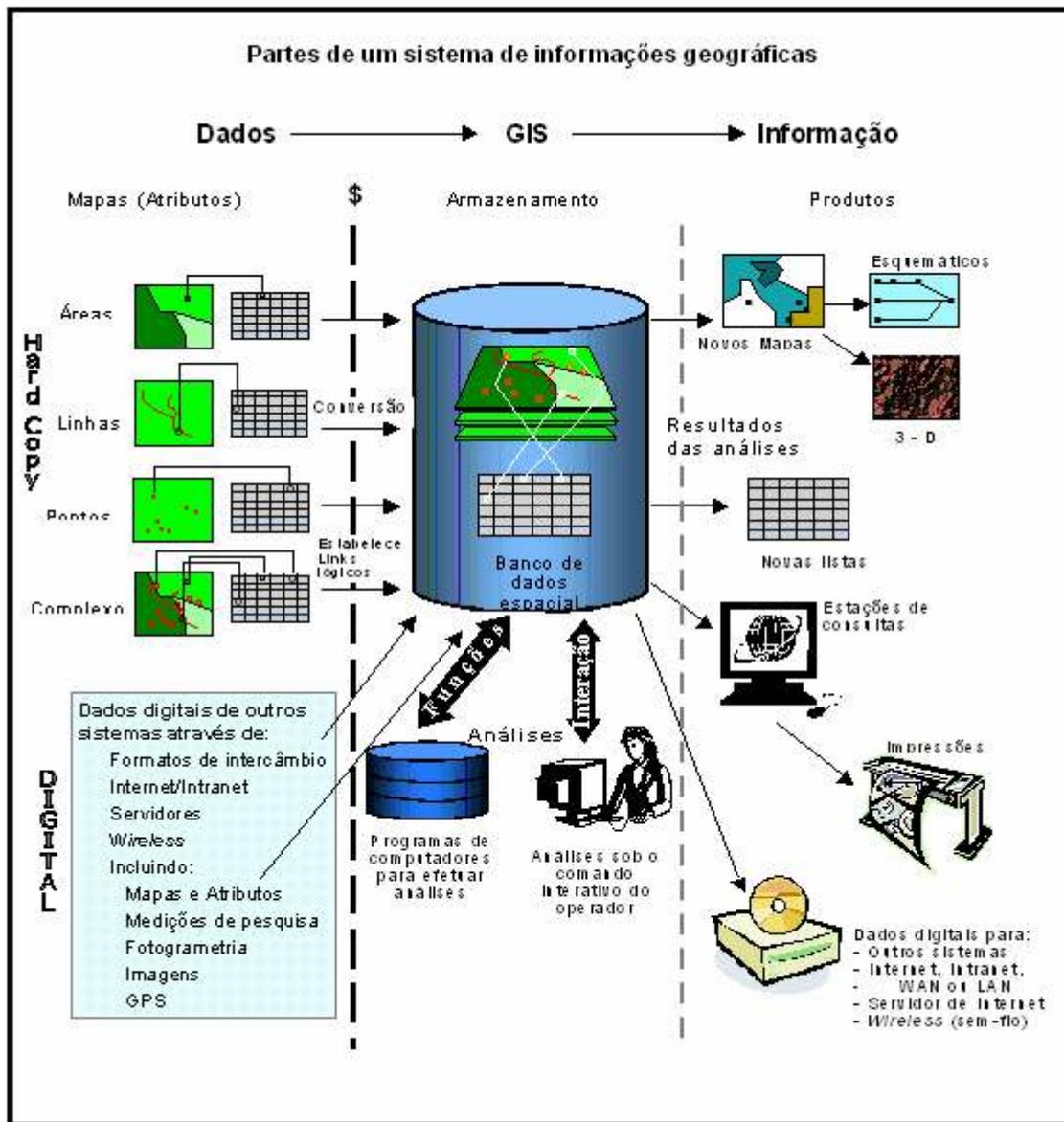


Figura 2.1.1 – Modelo holístico e funcional de um sistema de informações geográficas ¹⁸.

¹⁸ Adaptado de TOMLINSON, Roger F., "Thinking About GIS: geographic information system planning for managers" Readlands, CA: ESRI, 2003, pag. 3.

2.1.2 – Componentes de um GIS

Um componente básico do GIS e sobre o qual todos os demais, de uma forma ou de outra, são dependentes é o chamado *dado espacial*. O dado espacial é caracterizado pela presença da associação geográfica, ou seja, o dado conectado a um lugar conhecido na Terra, uma referência geográfica verdadeira. As diversas feições que normalmente encontramos nos mapas, tais como rodovias, lagos, edificações, entre outras, são normalmente encontradas no banco de dados GIS como camadas espaciais individuais. A maioria pode ser representada como uma combinação de pontos, linhas e polígonos. Associados às feições, encontramos os dados sobre cada uma, ou seja, os dados sobre tais rodovias, lagos e edificações. Quanto mais rico é o conjunto de dados que dispomos sobre uma feição, mais poderoso o dado espacial se torna no contexto do GIS.

Os dados que alimentam o GIS são obtidos de diversas maneiras e se encontram em várias formas e formatos. Uma fonte muito utilizada para a entrada de dados no GIS são os mapas impressos em papel. O mapa em papel é, na verdade, um banco de dados espacial e, ainda hoje, uma ferramenta gráfica muito utilizada pelo homem para o armazenamento, consulta e análise de dados espaciais. Com a evolução da ciência da computação e o surgimento do GIS, novas formas de armazenamento, de consulta e de análise foram desenvolvidas e os mapas se tornaram uma fonte natural e importante de obtenção de dados espaciais para armazenamento em GIS, através dos processos de escaneamento (*scanning*) ou digitalização.

Muitos dados referentes às feições espaciais nas organizações estão armazenados em listas e tabelas. É necessário que se estabeleçam associações lógicas para que tais dados possam ser posteriormente convertidos e utilizados no GIS.

Os dados também são cada vez mais freqüentemente encontrados na forma digital. Estes podem ser adquiridos por fotogrametria, GPS, imagens orbitais e instrumentos de medição que geram terabytes de dados úteis ao GIS. Também podem ser comprados ou obtidos gratuitamente pela internet ou através de convênios entre as empresas. Todos estes dados são integrados via associações lógicas dentro do banco de dados GIS sobre uma chave de organização primária que é a localização geográfica.

Um *GIS associa os dados espaciais com as informações geográficas de dada feição. A*

informação é armazenada como *atributo* da feição geograficamente representada. Por exemplo, um ponto que representa um imóvel não fornece nenhuma informação além da sua localização espacial. Para encontrar o número, o nome do proprietário, o consumo médio mensal de água do imóvel, é necessário executar uma consulta à base de dados. Utilizando a informação armazenada, é possível criar uma consulta que simbolize o imóvel em função do tipo de informação que precisa ser exibida. Por exemplo, uma consulta poderia ser efetuada, para exibir apenas os imóveis que possuem um consumo médio mensal de água acima de 50 m³.

Todos estes dados logicamente associados residem em um sistema de armazenamento do GIS que os torna disponíveis para consultas, pesquisas, análises, comparações e medições. Diversas funções de análise são efetuadas interativamente por um ou vários operadores cujo trabalho é criar as informações necessárias para a tomada de decisão. Estes produtos podem variar, dependendo do problema ou da necessidade, e incluem novos dados espaciais, novos atributos, novas listas e tabelas, resultados de pesquisas e consultas em vídeo, relatórios e mapas em papel, além de informações digitais.

2.1.3 – Abrangência do GIS

Um aspecto importante a ser considerado no GIS diz respeito à sua abrangência ou ao seu contexto de implementação. Talvez seja esta uma das fontes de confusão a respeito do GIS e do seu significado.

Um GIS pode ser implementado em um projeto localizado, ou seja, seu poder de manipulação de dados espaciais e de análise é utilizado para a solução de um caso específico e bem determinado. Nesta circunstância, o GIS é normalmente utilizado para a tomada de decisão a respeito de um projeto específico na produção de um determinado resultado, sendo que sua utilização termina quando o resultado é obtido.

Outra forma de implementação do GIS é a construção de um aplicativo departamental, onde uma importante função da organização é atendida por um sistema estruturado. O departamento responsável por esta função opera este sistema cujo objetivo é produzir uma direção estratégica duradoura para um determinado negócio da corporação, embora tenha seus produtos restritos ao departamento.

Uma terceira forma de implementação é o chamado GIS corporativo que permite a todos os membros da organização acessar e integrar dados do GIS através de todos os departamentos. O GIS corporativo suporta as necessidades dos departamentos e as decisões estratégicas de negócios de múltiplos departamentos. Neste cenário, o GIS se torna uma poderosa ferramenta dentro da organização e alinha-se com os seus objetivos estratégicos, passando a apoiar as decisões estratégicas de negócio de toda a organização.

Graças à Internet, o GIS está evoluindo da corporação para a sociedade. Alguns observadores da indústria prevêm que o próximo passo é a implementação do GIS em um contexto social amplo que rompe as fronteiras da corporação onde cada parte funcional da sociedade perceberá que eles fazem uso do GIS da mesma maneira como hoje eles fazem uso dos computadores (TOMLINSON, 2003).¹⁹ Os primeiros passos já foram dados. Podemos observar a proliferação de serviços na internet²⁰ oferecendo funcionalidades com o uso do GIS, desde a produção de mapas de localização nas mais diversas escalas, até a elaboração de mapas interativos informando rotas de percurso, opções de lazer, dados censitários, informações climáticas, entre outras informações. Também o uso da análise espacial e das propriedades dos objetos espaciais já pode ser observado.

¹⁹ TOMLINSON, Roger F., "Thinking About GIS: geographic information system planning for managers" Redlands, CA: ESRI, 2003, pag. 6.

²⁰ Alguns exemplos podem ser encontrados nos seguintes endereços da Internet: < <http://mapas.folha.com.br/>>, < <http://mapas.folha.com.br/>>, < <http://www.mapquest.com/>>, < <http://geohosting.geoexplore.com.br/website/copasa>>, < <http://plasma.nationalgeographic.com>>, acesso em 10 ago. 2005.

2.1.4 – Camadas espaciais

A organização da base de dados espaciais no banco de dados GIS requer freqüentemente uma estruturação do objeto ou fato geográfico em camadas ou *layers* que contém uma ou mais feições espaciais de pontos, linhas e polígonos.

Uma camada agrupa feições similares em uma base de dados espacial. Cada camada representa um assunto, tema ou classificação separada do mesmo espaço (figura 2.1.2). Para o ambiente de saneamento, por exemplo, poderiam ser redes de água, redes de esgoto, zonas de abastecimento, clientes e assim por diante.

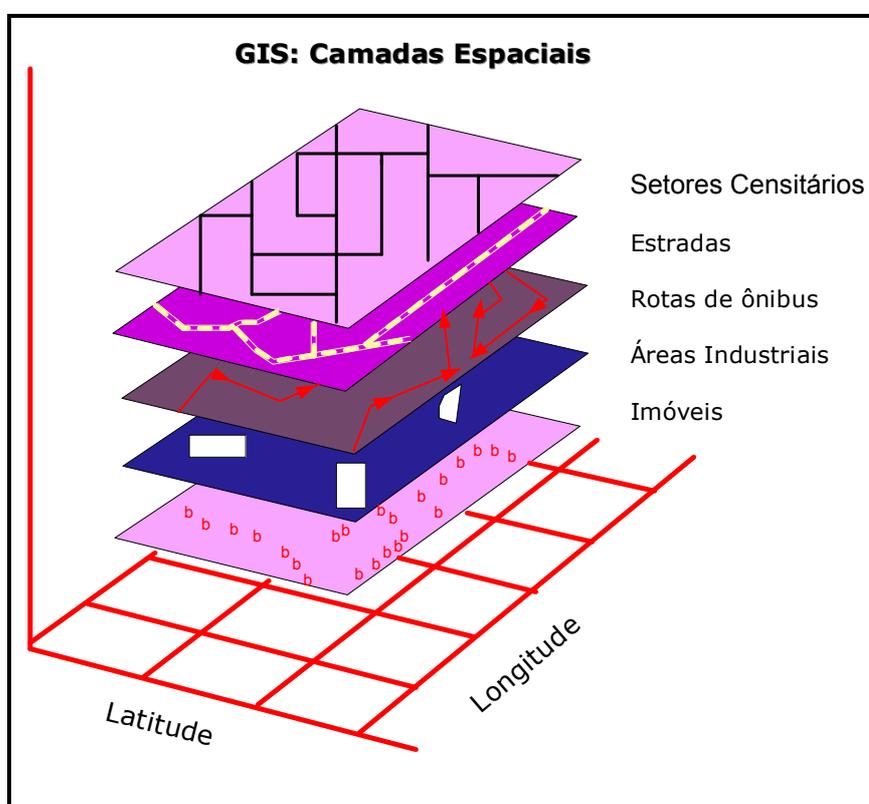


Figura 2.1.2 – Camadas espaciais do GIS ²¹

Em várias ocasiões, surge ainda a necessidade de se trabalhar com diversas camadas ao mesmo tempo. Muitas vezes é necessário combiná-las ou sobrepor umas às outras, para a obtenção de um resultado final. Esta combinação ou *overlay* de camadas é também uma característica importante nos ambientes *GIS*.

²¹ Adaptado de: FOOTE, K.E., LYNCH, M. "GIS as an Integrating Technology: Contexts, Concepts, and Definitions", Department of Geography, University of Texas at Austin, 1997, (http://www.colorado.edu/geography/gcraft/notes/intro/intro_f.htm).

2.1.5 – Objetos Espaciais

Uma das formas de se representar os objetos espaciais em GIS é através de um modelo de feições. Este modelo divide o mapa em três diferentes tipos de feições: pontos, linhas e polígonos. Um GIS com estrutura topológica constrói feições complexas de uma forma hierárquica utilizando estas três feições básicas (figura 2.1.3).

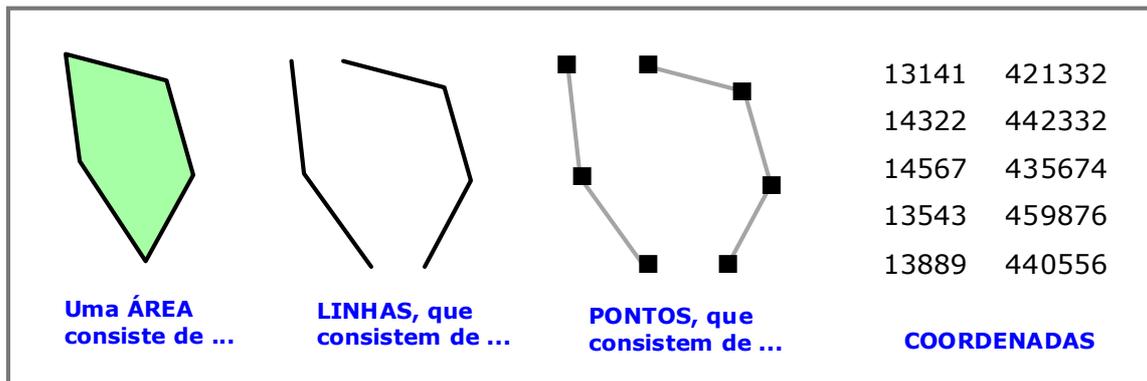


Figura 2.1.3 – Hierarquia de construção de feições complexas.

Pontos conectados formam linhas, linhas conectadas formam áreas. Estas feições complexas do GIS produzem as seguintes características:

- Volume
- Dimensionalidade (pontos possuem dimensão zero, linhas são unidimensionais e áreas são bidimensionais).
- Continuidade/autocorrelação espacial (feições de um local particular possuem maior tendência de serem semelhantes a aquelas de um local próximo, com menos similaridade a distâncias maiores).

2.1.6 – Propriedades das feições

Uma compreensão das propriedades dos objetos geográficos permite um processo de geocodificação mais efetivo, provê um uso correto das estruturas de dados cartográficos, e facilita o uso das transformações do dado cartográfico. As propriedades são as seguintes (figura 2.1.4):²²

²² CLARK, Keith C., "Analytical and computer cartography", 1995, Prentice-Hall, Inc., 2nd Edition, pp 49-80.

- Tamanho
- Distribuição
- Padrão
- Contigüidade
- Vizinhaça
- Forma
- Escala
- Orientação

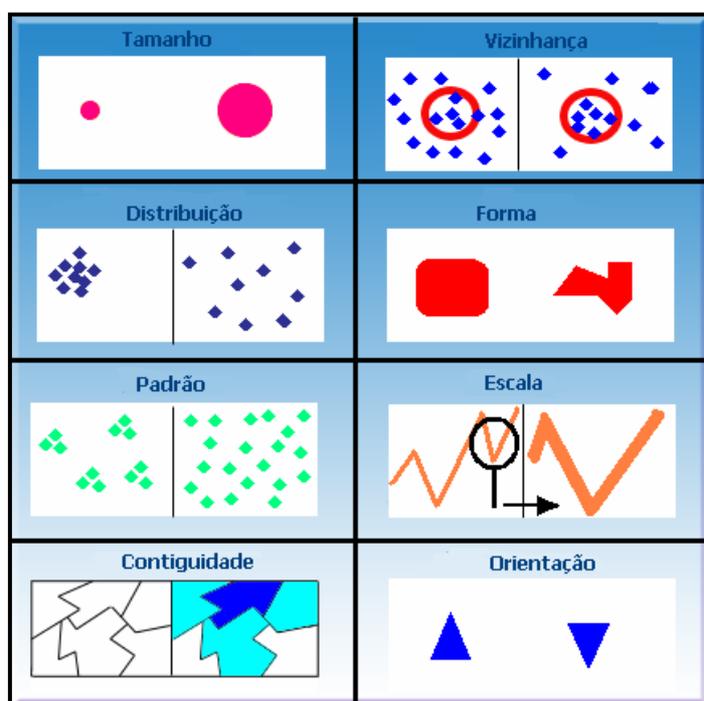


Figura 2.1.4 – Propriedades das feições dos objetos geográficos²³

O tamanho é uma propriedade básica e é caracterizado pela medição. A maioria dos fenômenos geográficos pode ser medida diretamente como em uma foto aérea. Um ponto é medido pelos seus aspectos de localização (x,y), adjacência e elevação. Uma linha possui comprimento, direção e conectividade. Um polígono possui topologia, área, forma, comprimento de limites, bem como localização e orientação. Um volume possui topologia, continuidade, inclinação de superfície, aspecto da superfície, tendência da superfície, estrutura, localização e elevação.

²³ Adaptado de: CLARK, Keith C., "Analytical and computer cartography", 1995, Prentice-Hall, Inc., 2nd Edition, pag 53.

A distribuição é medida pela densidade do fenômeno no espaço. A densidade pode ser calculada pela contagem dos objetos cartográficos ou de seus atributos em um determinado conjunto de unidades cartográficas, tais como um *grid* ou um conjunto de regiões. A densidade de um fenômeno geográfico possui implicações não somente em como se pode medi-lo e geocodificá-lo, mas também em como generalizá-lo. A análise da distribuição do fenômeno em um mapa influencia nossa decisão e em como ele pode ser simbolizado.

O padrão é outra propriedade fundamental dos objetos geográficos. O padrão é, na verdade, uma característica da distribuição e é uma descrição de sua estrutura. O padrão pode ser considerado como uma falta de aleatoriedade, podendo ser medido, desde que se faça um modelo ou uma descrição do padrão que se deseja encontrar. A repetição é uma das formas mais simples de padrão juntamente com a orientação dos objetos distribuídos no espaço.

Se o padrão é a repetição de um atributo no espaço, a propriedade de vizinhança define como esta propriedade varia no espaço. Esta variação se relaciona com a distância. Assim, pequenas separações significam similaridade, grandes separações significam dissimilaridade. O relacionamento entre o fenômeno geográfico e a distância é frequentemente utilizado nos estudos geográficos que aplicam as funções matemáticas de autocorrelação, modelos de interação espacial, modelos gravitacionais, entre outros.

A contigüidade é a propriedade de estar relacionado por justaposição, ou seja, por dividir um limite comum. Contigüidade, portanto, é uma das expressões geográficas de topologia e pode ser expressa de muitas formas. Podemos defini-la em termos de fronteiras compartilhadas, que, neste caso, pode ser medida pelo número e comprimento dessas fronteiras. Em uma rede, a contigüidade é chamada de conectividade. Em uma estrutura de *grid*, a contigüidade é analisada pelos *pixels* contínuos a um *pixel* central chamado de *kernel*.

A forma é outra propriedade dos objetos geográficos e é muito difícil de ser medida. Assim, as medições de forma são, na verdade, medições do nível de correspondência entre formas. Por exemplo, pode-se medir a forma de um objeto comparando-a com um círculo de mesma área e sobrepor uma sobre a outra.

A última propriedade geográfica fundamental é a escala. A expressão mais simples da escala é a fração representativa, a razão das distâncias em um mapa para as mesmas distâncias do mundo que ele representa.

2.1.7 – Modelos de implementação de dados do GIS

Há maneiras diferentes de armazenar dados espaciais no computador, mas a maioria se resume nos modelos raster e vector. Ambos possuem seus pontos fortes e fracos um em relação ao outro, e ambos são importantes para a análise e consulta em GIS.

O modelo de dados raster é um modelo que matematicamente é representado por uma matriz onde linhas e colunas são usadas para criar um campo ou célula. A célula é onde qualquer linha e coluna se encontram (figura 2.1.5). Cada célula possui o mesmo comprimento e largura. O formato raster é o formato comum para os dados de imagens orbitais, das fotografias aéreas, dos mapas digitalizados em *scanners* e das imagens fotográficas em geral.

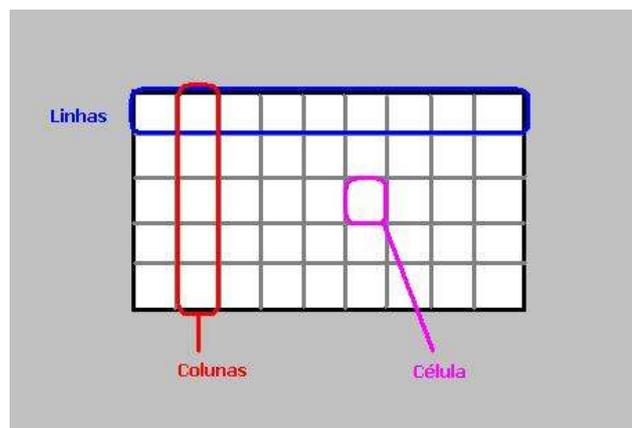


Figura 2.1.5 – Estrutura do grid no formato raster.

Algumas características deste modelo são:

- Uma célula somente pode conter um atributo, ao contrário da feição de ponto no formato vector que pode conter múltiplos atributos, tais como elevação, localização, tamanho, etc;
- Cada célula tem um valor mesmo que ele seja nulo;
- Uma célula pode conter um número ou um valor índice para um dado atributo, ou seja, índice de vegetação que representa uma percentagem de cobertura vegetal numa dada célula;
- Uma célula tem sua resolução dada de acordo com o tamanho da célula em unidades do solo, ou seja, um raster com uma resolução de 10 metros tem *pixels* de 10 metros

de lado, ou 100 metros quadrados.

Entre as vantagens do modelo raster podemos citar: os dados raster possuem uma estrutura simples e as operações de superposição são facilmente e eficazmente implementadas. Também altas variabilidades espaciais são eficazmente representadas; operações de modelagem e simulação são facilitadas porque cada unidade espacial tem a mesma forma e tamanho. Porém, este modelo apresenta algumas desvantagens: a estrutura dos dados ocupa muito espaço de memória; as relações topológicas são difíceis de serem representadas; transformações de projeção consomem muito tempo; o uso de grandes pixels, para reduzir o volume de dados, pode significar que estruturas fenomenologicamente reconhecíveis podem ser perdidas.

Os modelos vetoriais são baseados em elementos geométricos básicos - chamados de primitivas (linha, círculo, polígono, arco...), situados num sistema de coordenadas bi ou tridimensional, e produzidos por meio de sua descrição matemática. As primitivas geométricas servem de base para a criação/composição de modelos mais complexos e elaborados e, uma vez possuindo seus pontos referenciados em X, Y e Z (ainda que o Z seja zero, no caso dos modelos em duas dimensões), para isso podem ser manipuladas por meio de operações de transformação. No modelo vetorial pontos são armazenados por sua coordenada real na Terra; linhas e áreas são construídas por seqüência de pontos.

Uma variação do modelo vetorial é o modelo arco-nó que utiliza pontos na forma de nós e vértices para formar uma linha (figura 2.1.6).

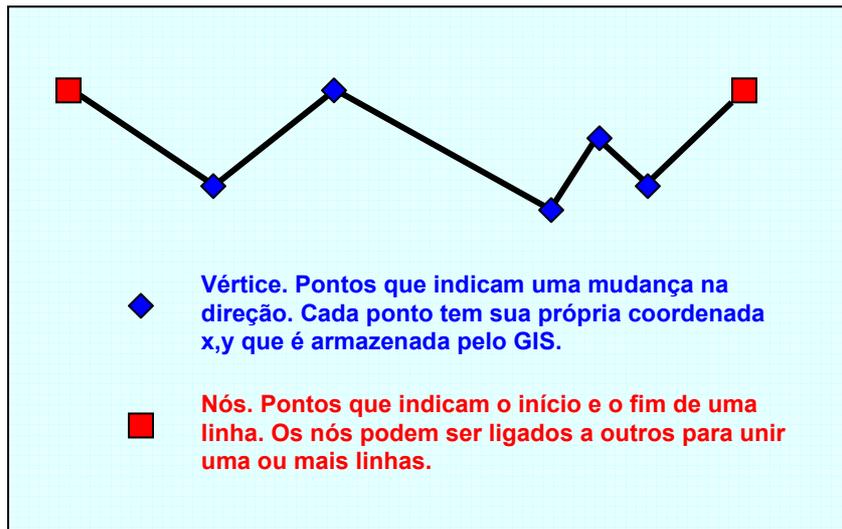


Figura 2.1.6 – Estrutura do modelo vetorial arco-nó.

Um ponto é um simples par de coordenadas (x,y) representando uma característica geográfica para a qual não se interessa ter uma área definida. Um ponto não tem propriedades geométricas além da localização, que é definida pelas coordenadas x e y.

Normalmente nos ambientes *GIS*, um ponto é numerado e armazenado com suas coordenadas e uma série de atributos. Os atributos descrevem as características da feição de ponto (figura 2.1.7).

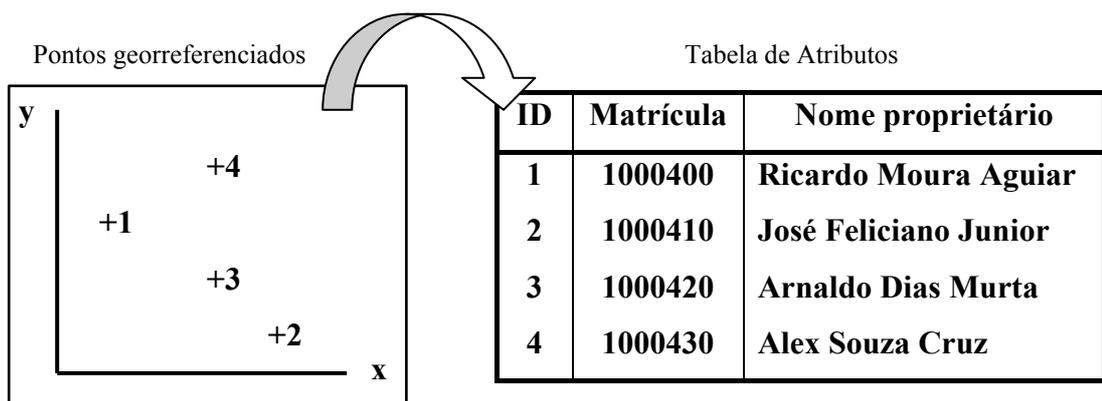


Figura 2.1.7 – Atributos associados a pontos georreferenciados.

Um arco representa a característica linear em uma base de dados espacial. Um arco tem as propriedades geométricas de comprimento e localização, definidas por seu conjunto de coordenadas x, y. O arco possui os nós de início e fim, além das coordenadas dos vértices quando há uma variação qualquer na direção do mesmo.

Geralmente na estrutura de armazenamento do GIS, um arco é numerado e armazenado com seu conjunto de pares de coordenadas e atributos (figura 2.1.8). Esses atributos podem incluir o comprimento (*length*) do arco e suas relações com outros arcos e polígonos.

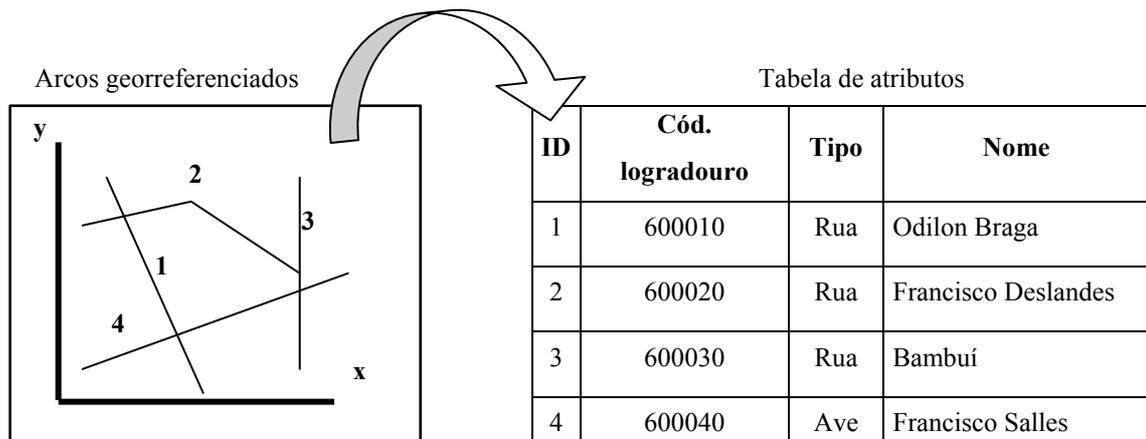


Figura 2.1.8 – Atributos associados a arcos georreferenciados.

Polígonos representam feições de mapas que têm as propriedades geométricas de área e perímetro (figura 2.1.9). Um polígono é composto de arcos compreendendo uma área, e um ponto de identificação dentro da área, denominado *centróide*. Em alguns ambientes GIS, em uma camada de polígonos encontramos os três tipos de feições: pontos, arcos e polígonos.

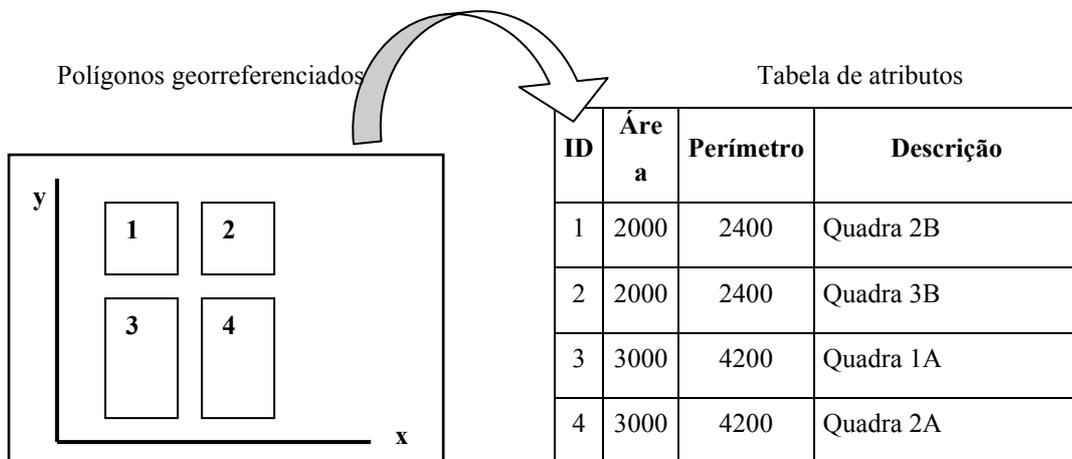


Figura 2.1.9 – Atributos associados a polígonos georreferenciados.

A seleção do modelo de implementação de dados a ser usado depende do fenômeno a ser estudado e das análises que se fazem necessárias, por exemplo: feições contínuas como temperatura e precipitação são melhor representadas no modelo raster, enquanto objetos discretos como pontos e linhas são mais adequados ao modelo vetorial.

2.1.8 – Topologia em GIS

A topologia em GIS pode ser definida como a conexão entre linhas, o conhecimento das fronteiras das áreas, e a inclusão de pontos em áreas. Em outras palavras, topologia é o relacionamento entre os três tipos de feições de mapas: áreas, linhas e pontos.

A topologia estuda importantes relações espaciais, tais como: quais polígonos estão localizados à esquerda e à direita de um arco; a direção e a conectividade do arco; a localização de pontos dentro de polígonos.

A topologia é armazenada com o arco (figura 2.1.10) e serve para dar ao arco uma “identidade espacial”, ou seja, como o arco está relacionado com todas as outras feições ao seu redor. A topologia também auxilia na detecção e eliminação automática de erros dos processos de geocodificação. Assim, qualquer linha desconectada ou qualquer polígono aberto pode ser detectado e, em algumas situações, corrigido automaticamente.²⁴

²⁴ Um exemplo de um verdadeiro modelo vetorial topológico pode ser encontrado nos arquivos TIGER – Topologically Integrated Geographically Encoded Reference – mantido pelo departamento de censo norte-americano e feito estritamente para o GIS. Sua estrutura contém feições geográficas como rios, lagos, estradas, fronteiras políticas e as informações censitárias para todo os Estados Unidos. Também as informações como latitude e longitude, tipo de feição, e endereços estão incluídas.

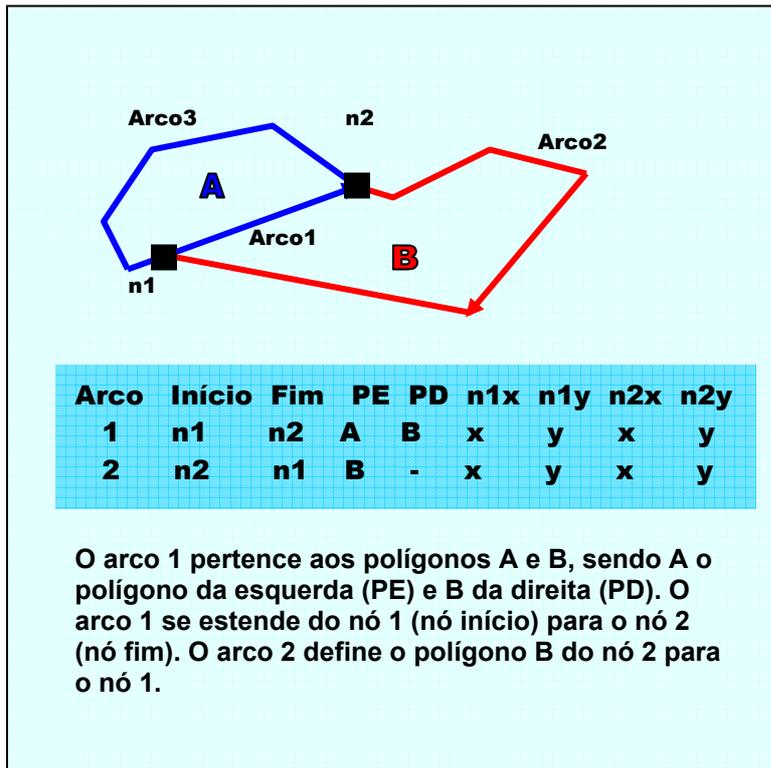


Figura 2.1.10 – Topologia básica do arco

2.1.9 – Qualidade dos dados espaciais ²⁵

Um dos fundamentos de um bom sistema de informações geográficas, bem como de qualquer outro sistema de informações, é a utilização de dados com qualidade. Ainda hoje, muitos sistemas de informações contêm uma quantidade significativa de dados imprecisos e de baixa qualidade.

A falta de dados com qualidade provoca custos dentro das organizações que são computados em atividades de correção, na perda de clientes, na perda de oportunidades de negócio e em decisões erradas. Em muitos projetos a qualidade dos dados geográficos não é examinada com atenção até que se descubra que alguma decisão errada foi tomada devido à sua baixa qualidade. Em outras palavras, usar um dado errado pode gerar um custo não desejado. Entretanto, dados precisos requerem uma atenção cuidadosa no planejamento e projeto de sistemas, no controle constante da captura dos dados e em ações agressivas de correção dos problemas que geram ou propagam dados imprecisos.

²⁵ ARONOFF, Stan, "Geographic Information Systems: A Management Perspective", 3rd Edition, 1993, WDL Publications.

Um dado possui qualidade se ele satisfaz os requisitos do uso que dele se deseja fazer (OLSON, 2003).²⁶ Assim, um dado é dito sem qualidade se ele não satisfaz estes requisitos. Em outras palavras, a qualidade do dado depende tanto do uso que dele se pretende fazer como do dado em si mesmo. Para satisfazer o uso pretendido, o dado deve ser preciso, ser fornecido a tempo, ser completo, ser compreensível e confiável. Entretanto, dependendo do uso, um determinado dado pode ser considerado de boa qualidade para uma aplicação e, ao mesmo tempo, de má qualidade para outra aplicação.

Para uma melhor compreensão dos aspectos referentes à qualidade do dado, deve-se diferenciar acurácia e precisão. Acurácia refere-se à proximidade entre o valor obtido experimentalmente e o valor verdadeiro (ou valor aceito como verdadeiro) na medição de uma grandeza física. Como o dado espacial é normalmente uma generalização do mundo real, é geralmente muito difícil identificar um valor verdadeiro, e trabalha-se com valores aceitos como verdadeiros. Por exemplo, na medição da acurácia do contorno em um banco de dados digital, compara-se o contorno com o desenhado em um mapa fonte, uma vez que o contorno não existe como uma linha real na superfície da Terra.

A precisão refere-se ao número de casas decimais ou dígitos significativos em uma medição. Precisão não é o mesmo que acurácia, ou seja, um grande número de dígitos significativos necessariamente não indica que a medição é exata ou acurada.

- **Componentes da qualidade dos dados espaciais**

Alguns componentes do dado devem ser avaliados para determinar sua qualidade. Estes componentes podem ser agrupados em um nível individual (fatores que fazem referência a cada elemento individualmente), em um nível mais global (diz respeito ao conjunto dos dados) e em um nível referente à sua utilização.

No nível individual podemos incluir a **acurácia posicional**, a **acurácia dos atributos**, a **consistência lógica** e a **resolução**. Estes componentes são usualmente testados estatisticamente.

²⁶ OLSON, Jack E., "Data Quality: The Accuracy Dimension", Morgan Kaufmann Publishers, San Francisco, 2003, pag. 24-27.

A **acurácia posicional** é a medida de quão próxima a informação de localização (normalmente as coordenadas) está da sua real posição na Terra. Ela é normalmente estimada calculando-se o desvio padrão dos pontos testados. Um desvio padrão baixo indica que a dispersão de erros posicionais é pequena.

Convencionalmente, os mapas possuem uma acurácia da largura de uma linha ou 0.5 mm, equivalente a 12 metros em uma escala de 1:24.000 ou 125 metros em mapas de 1:250.000. Tomemos o exemplo de um par de coordenadas UTM 579124,349 Leste e 5194732,247 Norte. Se o banco de dados foi digitalizado a partir de uma folha na escala 1:24.000, os últimos quatro dígitos em cada coordenada (unidades, décimos, centésimos e milésimos) seriam espúrios.

A **acurácia dos atributos** é a medida de quão próximos os valores dos atributos estão de seus valores verdadeiros. A acurácia dos atributos devem ser analisadas de diferentes maneiras dependendo da natureza do dado. Para atributos contínuos (superfícies) tais como em um modelo de elevação de terreno (DEM), a acurácia é expressa como erro de medição, por exemplo, uma elevação com acurácia de 1 metro.

Para atributos discretos, tal como em uma classificação de polígonos, podemos analisar:

- As categorias são apropriadas, suficientemente detalhadas e definidas?
- Existem erros de classificação, tais como: um polígono classificado como “A” quando deveria ser classificado como “B”?
- A heterogeneidade dos polígonos. Por exemplo, zonas de vegetação nas quais uma área pode ser 70% de uma determinada região “A” e 30% de uma determinada região “B”;
- Indefinição quanto à classificação, ou seja, “A” e “B” podem não definir claramente uma área, como tipicamente acontece na classificação de solos;

A **consistência lógica** diz respeito a quão lógicos estão os níveis de informação. Por exemplo, não há lógica encontrar um poço artesianos no meio de uma linha de centro de uma avenida em Belo Horizonte. Não há medida padrão para a consistência lógica, a qual deve ser identificada nos primeiros estágios da obtenção dos dados.

A **resolução** é a menor unidade que pode ser mapeada. Em uma imagem de satélite a resolução refere-se à dimensão do menor objeto que pode ser visto, assim chamada de

resolução espacial. Para os sistemas de escaneamento digital, tal como os sensores de satélites, a resolução espacial é definida como o tamanho do pixel. Em mapas temáticos, o menor objeto que pode ser mapeado é conhecido como a Unidade Mínima de Mapeamento (UMM).

Em um nível global, os componentes que determinam a qualidade dos dados são frequentemente avaliados por julgamento ou por relatórios de informações sobre os dados. São eles: **completeza**, **tempo** e **histórico**.

A **completeza** do dado refere-se à quantidade de dado disponível para uma determinada área de interesse. Pode acontecer que o banco de dados disponível não possua todos os dados necessários.

A dimensão do **tempo** pode ser crítica para muitos sistemas de informações geográficas. As feições em uma paisagem podem alterar-se muito com o passar do tempo. Aplicações em florestas podem necessitar de dados atualizados a cada 5 ou 10 anos, enquanto aplicações em agricultura podem necessitar de dados atualizados semanalmente. Dados mais antigos tendem a ser menos úteis que dados atualizados.

O **histórico** é um outro componente utilizado para uma avaliação da qualidade do dado. Através do histórico de produção de um dado e pelo exame da fonte do dado (informações normalmente obtidas dos metadados, ou dados sobre dados) pode-se ter uma melhor idéia da qualidade dos mesmos. Frequentemente o histórico é um indicador bastante útil da acurácia.

No que se refere à sua utilização dois componentes da qualidade do dado se destacam: sua **acessibilidade** e os **custos diretos** e **indiretos** de obtenção.

A **acessibilidade** se refere à facilidade de se acessar o dado. Em alguns casos a acessibilidade pode ser limitada quando o banco de dados é propriedade de um grupo privado.

Os **custos diretos** do banco de dados é o preço que se paga para utilizá-lo. O custo verdadeiro pode ser difícil de determinar se o dado é gerado no local, porque pode ser difícil estimar os custos do equipamento e serviços utilizados. Os **custos indiretos** dos bancos de dados são os custos resultantes de todo o tempo e material utilizado para se utilizar o banco de dados. Estes podem incluir os custos de digitalização ou simplesmente da familiarização

com o banco de dados que está em um formato diferente com o que se está familiarizado.

- **As fontes de erro em GIS**

Um outro aspecto que influencia na qualidade dos dados são os erros. Erros são comuns em todos os níveis da informação geográfica e são introduzidos a cada passo da manipulação e acesso aos dados geográficos. O erro, portanto, é muito difícil de ser eliminado totalmente. Entretanto, o nível de erro precisa ser controlado para que a informação não seja invalidada. A tabela 2.1.1 resume as fontes de erro mais comuns encontradas na utilização do GIS.

Estágio	Fontes de erro
Obtenção dos dados	Erros na entrada de dados de campos. Erros nos mapas existentes como fontes de dados. Erros na análise de dados obtidos por sensoriamento remoto.
Entrada dos dados	Imprecisões na digitalização em decorrência dos operadores e equipamentos. Imprecisões inerentes às feições geográficas (exemplo: limites de florestas que não são precisos).
Armazenamento dos dados	Precisão numérica insuficiente. Precisão espacial insuficiente.
Manipulação dos dados	Intervalos de classificação inapropriados. Propagação de erros na combinação de múltiplos níveis de informação (overlay). Fatiamentos causados por procedimentos de sobreposição de polígonos.
Saída dos dados	Imprecisões de escala. Erros causados por dispositivos de saída imprecisos. Erros causados pela instabilidade da mídia (exemplo: encolhimento do papel, etc.).
Utilização dos resultados	Erros de interpretação da informação produzida. Erros na utilização inapropriada da informação.

Tabela 2.1.1 – Fontes comuns de erros encontrados na utilização do GIS. ²⁷

²⁷ CLARKE, K. C., "Getting Started with Geographic Information Systems", Prentice Hall, 3rd. ed., 352 p., 2000.

2.2 – Geocodificação

2.2.1 – Introdução

A geocodificação, ou *geocoding*, é definida genericamente como a conversão da informação espacial para um formato digital, o que envolve a captura do mapa e, algumas vezes também, a captura dos atributos. Frequentemente, envolve a geocodificação de endereços, ou *address geocoding* (CLARK, 2000).²⁸ Portanto, os processos e os conceitos envolvidos na geocodificação determinam o tipo, a escala, a acurácia e a precisão de mapas digitais. Neste sentido, podemos considerar que a geocodificação é o primeiro estágio da cartografia computadorizada como também dos sistemas de informações geográficas.

O propósito da geocodificação é codificar digitalmente as características fundamentais do dado geográfico. Entre essas características, é fundamental para o dado geográfico o seu atributo de *localização* sobre a superfície da Terra. Embora nós tenhamos que usar uma terceira dimensão para descrever a elevação sobre a Terra, as duas dimensões de localização sobre o plano ou a esfera são provavelmente a propriedade geográfica básica. Normalmente, os valores x e y representam latitudes e longitudes, mas frequentemente na geocodificação utiliza-se uma projeção de mapa, como a *Transverse Mercator*, e um sistema de projeção tal como a *Universal Transverse Mercator* (UTM) para se fornecer localizações. A maioria dos sistemas de coordenadas para uso em computadores é baseada em coordenadas cartesianas. Isto significa dizer que os eixos das duas direções são ortogonais, ou seja, eles formam um ângulo reto entre si. Isto nos permite especificar uma localização no espaço através de um par de coordenadas (x , y).

A geocodificação possui muitos objetivos conflitantes que devem ser considerados antes de se iniciar um processo de geocodificação. Como uma das maiores fontes de erro no processo de geocodificação é o erro humano e o custo da mão-de-obra é normalmente alto, um objetivo importante da geocodificação é a busca da redução do trabalho manual envolvido. Também se deve buscar a detecção e eliminação de erros, ou seja, todo esforço deve ser feito no processo de geocodificação para evitar ou reduzir erros, ou mesmo para detectar mais facilmente o erro.

²⁸ CLARKE, K. C., "Getting Started with Geographic Information Systems", Prentice Hall, 3rd. ed., 352 p., 2000.

Outros objetivos também podem ser assinalados, como a otimização do armazenamento de dados, ou seja, a geocodificação freqüentemente busca eliminar informações redundantes. Por fim, um objetivo crítico é a maximização da flexibilidade, que busca estruturas de dados e métodos de armazenamento que permitam análises não previstas anteriormente. Tão logo o dado venha a existir em uma forma digital, novas aplicações para eles serão descobertas e é importante que estes não contenham restrições na sua acurácia, precisão e confiabilidade.

- **A geocodificação com representação topológica**

A geocodificação pode ser realizada em dois níveis diferentes. Pode-se simplesmente converter os elementos gráficos de um mapa em números de forma que o mapa possa ser posteriormente reproduzido usando os métodos da cartografia computadorizada, ou pode-se também codificar informações topológicas importantes sobre o dado que se quer geocodificar. Assim, se o mapa consiste de uma estrada e um rio que se cruzam, uma abordagem não topológica seria a representação das duas linhas uma sobre a outra simplesmente. Mas uma representação topológica poderia reconhecer que ou a estrada cruza o rio sobre uma ponte, ou o rio cruza a estrada por um aqueduto ou canal.

Um exemplo da aplicação da geocodificação topológica pode ser observado nos esforços do escritório norte-americano para o censo ou, o assim chamado, *U.S. Census Bureau*, em encontrar formas de mapear os dados reunidos em todo os Estados Unidos, endereço por endereço. No Censo para População e Habitação de 1960, questionários foram enviados pelo correio e recolhidos de cada lar por funcionários dos correios. Em 1970, o plano era usar os correios para a entrega e a devolução dos formulários de pesquisa. Isto demandou a capacidade de geocodificação e, subseqüentemente, o desenvolvimento de um guia de código de endereçamento (*Address Coding Guide – ACG*).

Os primeiros esforços de geocodificação permitiram somente a digitalização de logradouros, mas logo a capacidade de mostrar quadras e setores censitários foi adicionada. Isto exigiu o reconhecimento das faces das quadras, o que foi feito pela digitalização dos nós representando as interseções. Isto, por sua vez, significou que as interseções tinham que ser numeradas e as faixas de endereços tinham que ser reconciliadas para as faces corretas da quadra. A forma das linhas no mapa tinha que ser precisamente determinada e anotada, criando a topologia do mapa. O nome dado a este novo processo de mapeamento de quadras foi em inglês chamado de *Dual Independent Map Encoding (DIME)*, ou codificação

independente dual de mapa, o que combinado com o processo de cruzamento de endereço, ou *address matching*, foi chamado de ACG/DIME. Em 1980, o padrão ACG/DIME se tornou o “arquivo geográfico básico” (*geographic base file – GBF/DIME*). A flexibilidade analítica foi a chave do sucesso dos arquivos DIME. Embora seus códigos geográficos tenham sido projetados como uma forma de se produzir mapas censitários, logo foi percebido que os arquivos DIME se prestavam a outras aplicações mais genéricas, tendo sido usados para a coleta de dados estatísticos, geração automática de centróides, mapeamento temático e *address matching* automático. Muito deste sucesso estava relacionado ao georreferenciamento dos dados censitários que estes arquivos possuíam, permitindo uma cartografia temática automatizada com os seus milhares de atributos disponíveis.

A isto se seguiu uma demanda nacional norte-americana, sem precedentes, para um mapa digital que viria a se chamar *TIGER*, ou seja, uma abreviatura para: *Topologically Integrated Geographic Encoding and Referencing*. Desta forma, o geógrafo Robert Max e sua equipe do escritório do censo norte-americano, implementaram o *TIGER* para o censo norte-americano de 1990.

Os arquivos *TIGER* contêm faixas de endereços em lugar dos endereços individuais. Uma faixa de endereço se refere à primeira e à última estrutura de numeração possível ao longo da face da quadra, mesmo que as estruturas físicas possam não existir (figura 2.2.1). Para cada cadeia de endereços entre o nó inicial e o nó final, há duas faixas de endereços, uma para números pares à direita e outra para números ímpares à esquerda.²⁹

²⁹ Uma explanação mais completa pode ser encontrada em U.S. Census Bureau (1997). 1997 *TIGER/Line® Files Technical Documentation*. Washington, DC: U.S. Census Bureau.

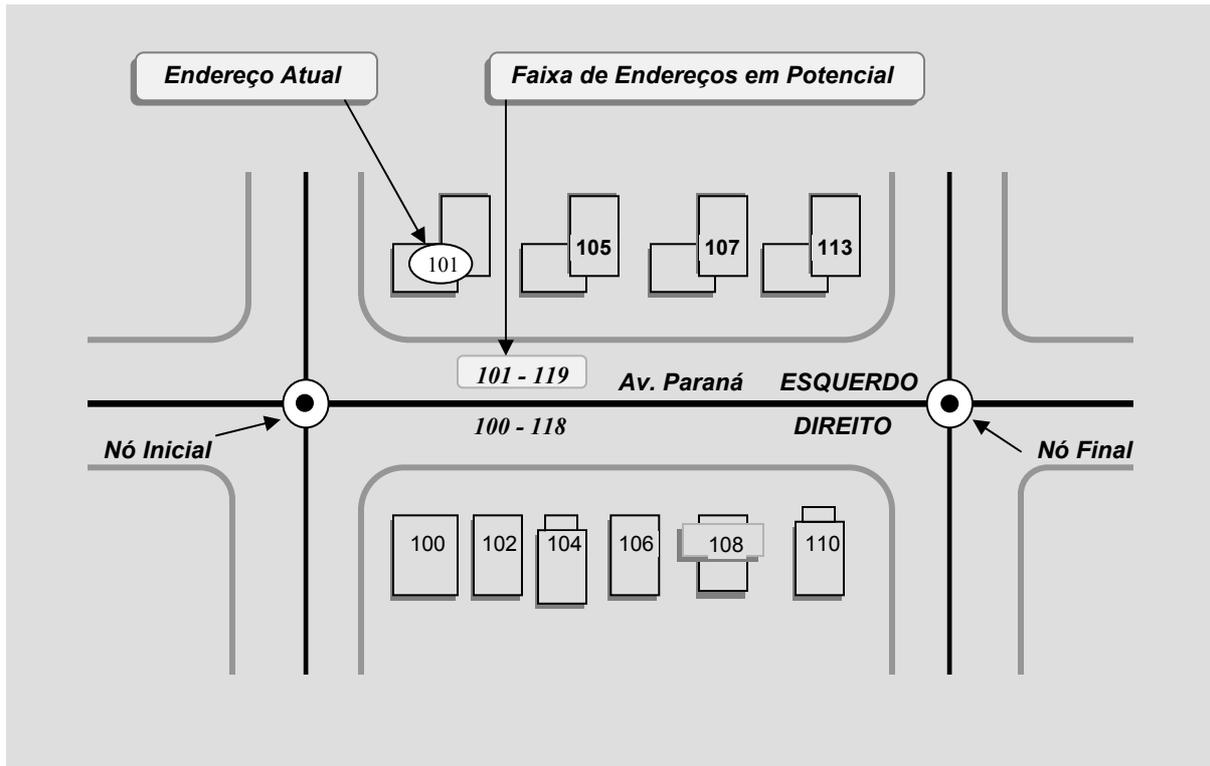


Figura 2.2.1 - Estrutura básica da faixa de endereços nos eixos de logradouros dos arquivos TIGER.³⁰

³⁰ Adaptado de: HARRIES, K., 1999, "Mapping Crime: Principles and Practice", (Washington: US Department of Justice).

2.2.2 – Geocodificação de endereços

O endereço pode ser considerado uma forma comum de informação de localização. Um endereço especifica uma localização basicamente da mesma forma que uma coordenada geográfica o faz. Mas os endereços são meramente caracteres de texto que contém um número de imóvel, um nome de logradouro e um código postal. Antes que se possa mostrá-lo em um mapa é necessário calcular a sua coordenada geográfica. A geocodificação do endereço (*address geocoding*) permite que se mostre dados de uma tabela que contenham endereços como pontos em um mapa. Para fazer isso, o processo de geocodificação associa endereços armazenados em um arquivo tabular com um arquivo com dados espaciais, freqüentemente com um arquivo de eixos de logradouros que também contenham os endereços. Dessa forma, as coordenadas das feições dos logradouros são utilizadas para se calcular e se obter as coordenadas dos endereços no arquivo. O resultado é um mapa no qual cada ponto representa um endereço do seu arquivo original de endereços.³¹

Há inúmeras aplicações para a geocodificação de endereços. Pode-se mapear os endereços de clientes, locais de emergência, membros de clubes, lojas de varejo, paradas de rotas de entrega, locais de crimes (assaltos, estupros, roubos, etc.), locais de doenças, entre outros. A habilidade de criar mapas a partir de arquivos de endereços e/ou outros locais geográficos é uma poderosa ferramenta para se explorar o potencial de informação do dado existente.

Geocodificação de endereços é a habilidade de adicionar pontos em um mapa a partir de um endereço de logradouro (ou outra informação de endereço). A geocodificação de endereços requer a comparação de cada endereço em um arquivo com as faixas de endereço no arquivo espacial de referência. Quando um endereço é achado dentro da faixa de endereços de um segmento de logradouro, uma interpolação é realizada para localizar e obter as coordenadas do endereço. Por exemplo, um arquivo contendo os endereços dos clientes pode ser combinado com um arquivo de logradouros. O resultado é um arquivo de pontos que mostra onde os clientes vivem. Os pontos resultantes seriam logicamente integrados com o banco de dados e utilizados como novas feições no banco de dados. As novas feições poderiam ser utilizadas por outras funções do sistema juntamente com o restante do banco de dados.³²

³¹ BRESLIN, Pat et al., "Getting to Know ArcView GIS", 1996, ESRI.

³² TOMLINSON, Roger, "Thinking About GIS: Geographic Information System Planning for Managers", ESRI, 2003, p268-269.

Uma das funcionalidades mais utilizadas e importantes do GIS é o processo de *address geocoding*, que busca transformar qualquer tipo de evento ou objeto localizado em um determinado ponto da superfície terrestre, em um objeto geográfico que pode ser representado em um mapa. Neste processo, é feita a identificação das coordenadas geográficas associadas aos endereços dos eventos ou objetos quaisquer de um banco de dados tabular. É uma poderosa tecnologia de GIS, porque ela pode transformar qualquer banco de dados que contenha endereços em um banco de dados GIS, que pode ser mapeado ou usado como matéria prima para análises de dados mais sofisticadas (Drummond, 1995).³³

Há um grande número de aplicações sócio-econômicas que requerem a manipulação pelo pesquisador de dados baseados em endereços em um sistema de informações geográficas. À medida que melhoram as capacidades do GIS e dos bancos de dados, um maior nível de resolução é possível e o endereço individual está se tornando um nível padrão para a investigação espacial.³⁴

Desde que surgiram, os sistemas de informações geográficas transformaram a forma como os planejadores urbanos conduzem as análises ambientais, registram os limites urbanos (quadras, lotes, vias, etc.), confeccionam os mapas de localização de infraestrutura, e modelam os sistemas de transporte urbano. Entretanto, esta tecnologia somente recentemente vem trazendo uma contribuição significativa na análise dos padrões da atividade humana que envolve os maiores problemas urbanos, incluindo temas como pobreza, crime, educação, violência, desemprego e saúde pública.

Embora o termo geocodificação esteja associado à utilização do computador, uma das primeiras aplicações de geocodificação de endereços de que se tem notícia foi levado a cabo pelo famoso epidemiologista inglês John Snow (1813-1858) em seu trabalho de pesquisa da epidemia de cólera em Londres no ano de 1854. Snow, nesta investigação, identificou manualmente a localização das mortes relacionadas à cólera em um mapa (figura 2.2.2), o que possibilitou uma melhor observação e compreensão do fenômeno.

³³ DRUMMOND, William J., "Address matching", Journal of the American Planning Association; Spring95, Vol. 61 Issue 2, p240, 11p.

³⁴ RATCLIFFE, Jerry H., "On the accuracy of TIGER-type geocoded address data in relation to cadastral and census aerial units"; International Journal of Geographical Information Science, 2001, Vol. 15, no. 5, 475-485.



Figura 2.2.2 – Casos de cólera mapeados pelo Dr. Snow em Londres em 1854.³⁵

³⁵ Porção do mapa original criado por Dr. John Snow. Através da identificação das mortes (uma linha paralela à face do imóvel em que as pessoas morreram) Dr. Snow foi capaz de verificar o crescimento do Cólera em torno da bomba de água na esquina das ruas Cambridge e Broad Street. Fonte: <http://www.csiss.org/classics/content/8>

2.2.3 – O processo de geocodificação de endereços

O objetivo da geocodificação de endereços é encontrar a localização em um mapa fornecendo um endereço. Um endereço encontrado resulta em uma coordenada geográfica. O processo de geocodificação de endereços (figura 2.2.3) pode ser definido como o processo de cruzamento de registros em dois bancos de dados baseado em um atributo de relacionamento comum, neste caso, o endereço. O primeiro destes bancos de dados, o banco de dados de referência, contém ambas as informações: o endereço e a localização baseada em um sistema de coordenadas geográficas.

Normalmente o banco de dados de referência contém um registro para cada segmento de logradouro, conforme exemplifica a estrutura dos arquivos TIGER ³⁶. Os arquivos de referência também podem consistir de dados detalhados de polígonos, tais como os limites de lotes, ou dados de pontos, tais como as localizações do código de endereçamento postal. Nesses casos, entretanto, não há a necessidade de se interpolar a informação de localização posicional, que é simplesmente copiada para o registro de endereço.

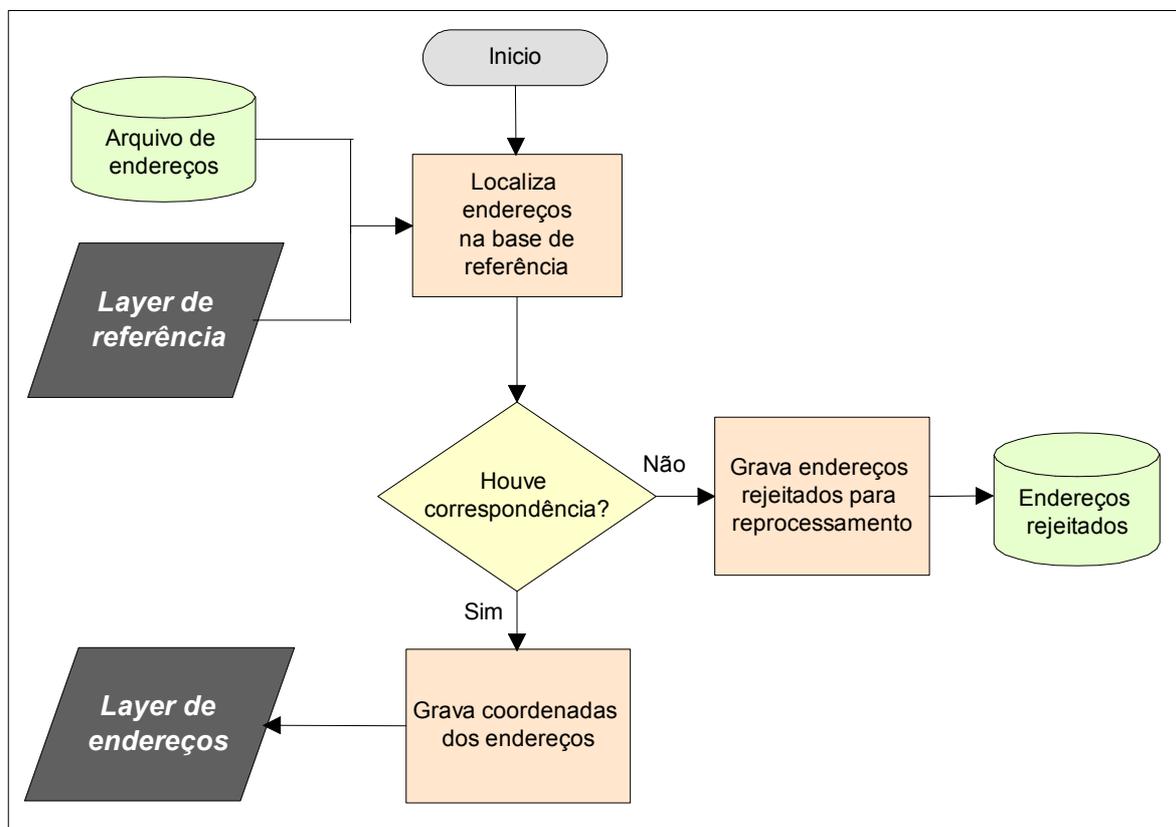


Figura 2.2.3 – Processo de geocodificação de endereços em uma base de referência de trechos de logradouros.

³⁶ Ver item 2.2.1.

O banco de dados de referência inclui campos como o nome do logradouro e o tipo do logradouro, além dos campos que armazenam a numeração máxima e mínima à direita e à esquerda do logradouro. Também são armazenadas as informações sobre a latitude e longitude dos nós de término dos trechos, bem como, opcionalmente, as informações de localização dos vértices intermediários entre os nós de término. Outros campos opcionais também podem armazenar o código de endereçamento postal, nome da localidade, código do setor censitário em cada lado do trecho do logradouro.

O segundo banco de dados requerido é o banco de dados de endereços alvo. Ele contém a informação do endereço mais outros campos adicionais que caracterizam o endereço ou o evento que teve lugar no endereço ou que está relacionado a ele. Os *softwares* de localização de endereço buscam identificar um segmento ou trecho de logradouro no banco de dados de referência que possui o mesmo nome e/ou tipo de logradouro, além de outros identificadores que possam existir em ambas as bases. Depois que os dois registros tenham sido correspondidos, uma interpolação é usada para associar uma coordenada geográfica ao registro do banco de dados alvo. Este processo de correspondência entre registros é um exemplo especial da maneira na qual um sistema gerenciador de banco de dados relacional pode unir dois bancos de dados que dividem um campo ou um conjunto de campos comuns.³⁷

Quando um registro do banco de dados alvo encontra um único registro no banco de dados referência onde o nome do logradouro, o tipo do logradouro, além de outros campos de ambos os registros, são idênticos, e o número do endereço do registro alvo está contido dentro dos limites de numeração do registro de referência, ocorre uma correspondência perfeita. Quando a maioria dos campos é idêntica, mas há uma diferença no campo “tipo de logradouro”, por exemplo, ocorre uma correspondência parcial que pode ou não ser aceita, dependendo do algoritmo de correspondência utilizado e do propósito final do processo.

Uma vez que o endereço alvo é correspondido a um único trecho de logradouro no banco de dados de referência, o algoritmo de correspondência verifica a paridade (designação par/impar) do endereço alvo e determina seu alinhamento à direita ou à esquerda do trecho baseado no sentido deste.

³⁷ Em um sistema gerenciador de banco de dados relacional as tabelas se relacionam através de um atributo comum frequentemente denominado de chave primária em uma tabela e secundária em outra.

O padrão estabelecido indica que a numeração par do logradouro deverá estar à direita do mesmo. Semelhantemente, a numeração ímpar deverá estar à sua esquerda (figura 2.2.4).

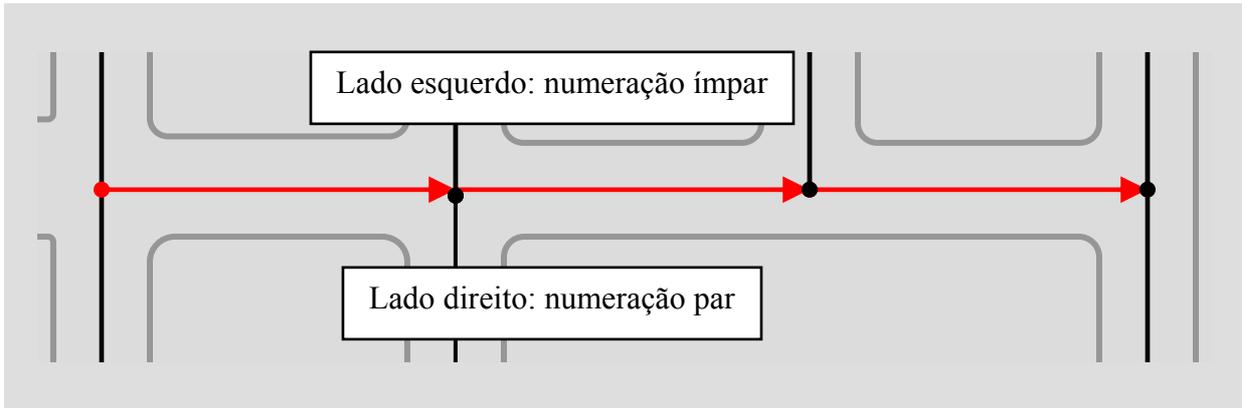


Figura 2.2.4 – Padrão de disposição da numeração ao longo do logradouro determinado pelo sentido do trecho.

O algoritmo, então, interpola para determinar a distância aproximada do endereço ao longo do trecho utilizando os valores de numeração máximo e mínimo da face da quadra (figura 2.2.5).

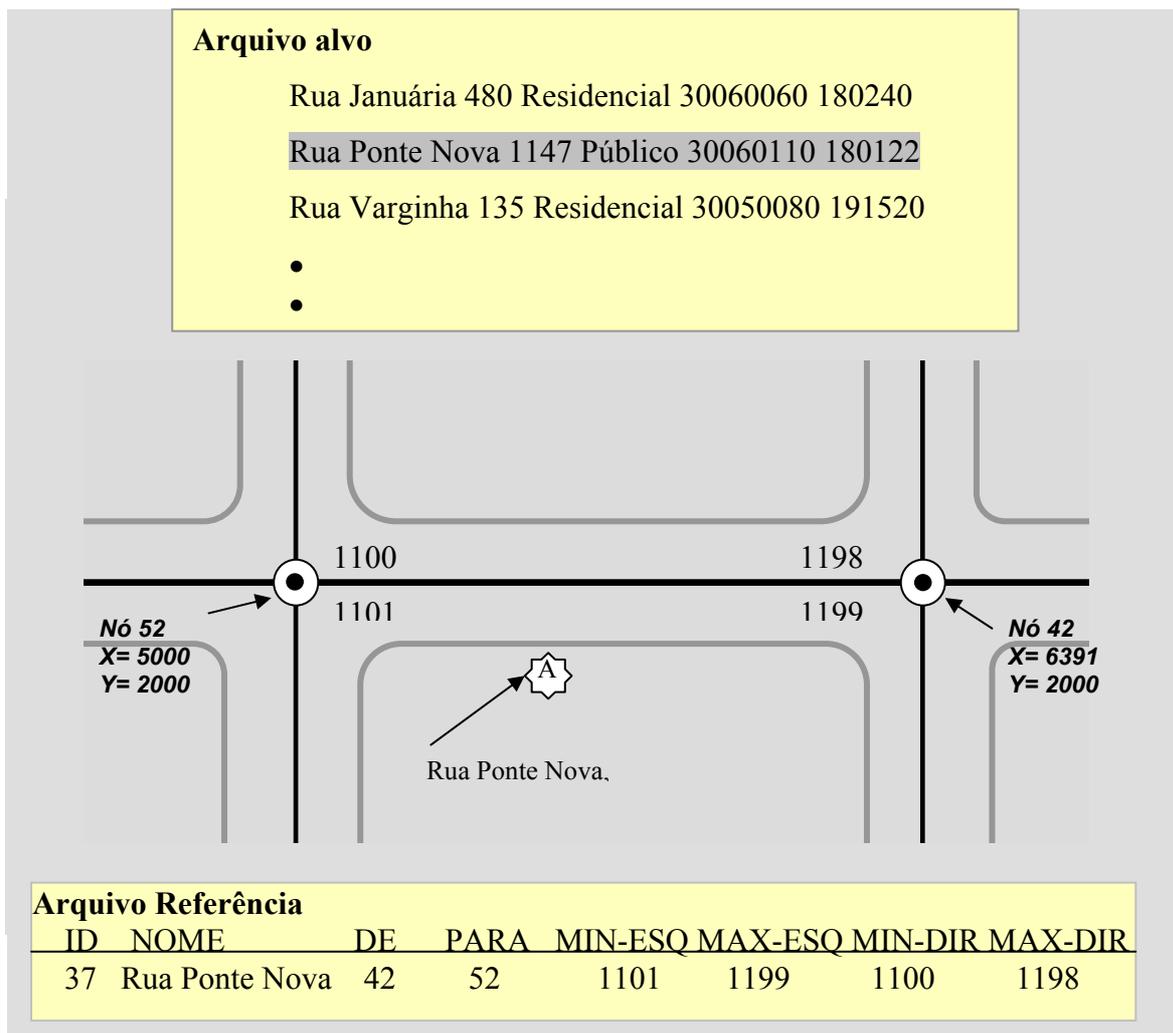


Figura 2.2.5 - Exemplo do processo de geocodificação de endereços em uma base de referência de trechos de

logradouros.

O exemplo da figura 2.2.5 tem o endereço A na base de endereços alvo cuja coordenada é encontrada, depois de identificado o trecho de logradouro correspondente no arquivo de referência (ID 37), através de interpolação:

Diferença do endereço de A na face: $1147 - 1101 = 46$

Percentual do endereço de A na face: $46 / (1199 - 1101) = 47\%$

Mudança da coordenada em X: $6391 - 5000 = 1391$

Coordenada em X interpolada do endereço A: $5000 + (0,47 * 1391) = 5653,77$

Neste exemplo, a coordenada em Y é constante. Assim, temos as coordenadas para o endereço A: (5653,77 ; 2000).

Em alguns *softwares* de geocodificação também são introduzidos alguns fatores de compensação para melhor ajustar as coordenadas dentro da quadra. Um deles é o fator de afastamento que especifica a distância de afastamento do ponto ao trecho (figura 2.2.6). Este fator pode ser especificado como uma constante, ou variar para cada trecho de logradouro de acordo com o conteúdo de um atributo do próprio trecho.

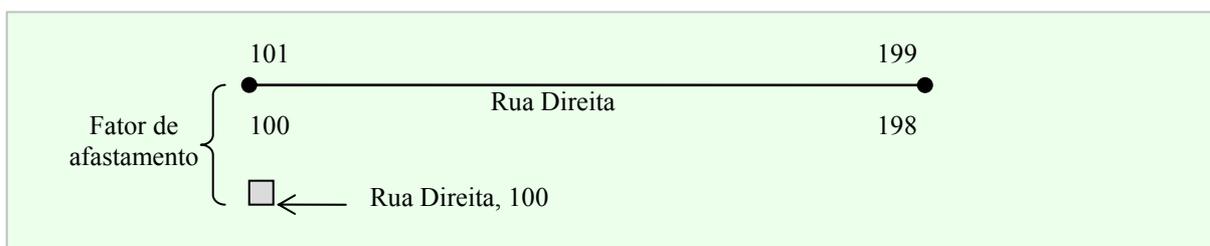


Figura 2.2.6 – Fator de afastamento na geocodificação de endereços.

Outro fator de ajuste é o fator de encurtamento usado para “puxar” os pontos georreferenciados para dentro do trecho. Este fator previne que o endereço de fim de trecho possa ser posicionado como se fosse localizado no logradouro adjacente (figura 2.2.7).

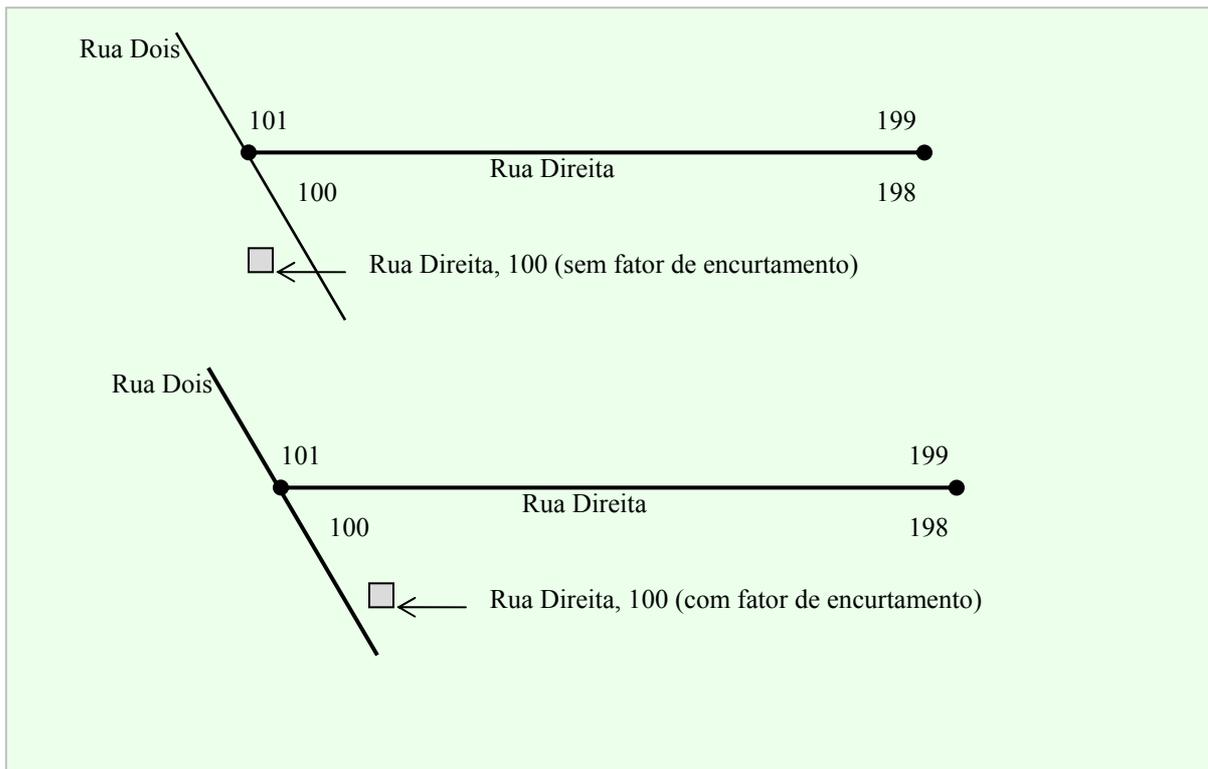


Figura 2.2.7 – Fator de encurtamento na geocodificação de endereços.

O fator de encurtamento é normalmente um percentual. Um valor de 10%, por exemplo, significa que o trecho será encurtado em 10% em seu comprimento antes de se realizar a interpolação do endereço e aplicado o fator de afastamento.

Finalmente, para que o processo de geocodificação apresente melhores resultados é normalmente necessário que os arquivos de referência e de endereços alvo sejam compatibilizados. Ambos os arquivos, infelizmente, podem armazenar informações sobre endereços em diferentes estilos, podendo haver divergências entre si e mesmo entre os próprios registros. Os números de imóveis, a abreviação de logradouros e os erros gramaticais são as fontes de inconsistência mais comuns.

Uma vez que os arquivos de referência e de endereços alvo tenham sido preparados, o processo de geocodificação pode iniciar. Tanto os pacotes de *software* específicos como as ferramentas de geocodificação do GIS permitem a realização do processo de duas formas: interativamente ou em lote. Durante o processo em lote, o aplicativo de geocodificação efetua a correspondência do maior número de endereços possível sem a intervenção do operador. Os endereços que não foram geocodificados, por alguma razão, podem ser geocodificados interativamente. A geocodificação interativa mostra cada registro não geocodificado para que

o operador possa editá-lo, a fim de corrigir os erros gramaticais, incluir alguma informação ausente, expandir alguma abreviação, ou excluir uma informação indevida que evitou uma correspondência perfeita. Os aplicativos de geocodificação também podem apresentar trechos de logradouros que são muito similares aos registros não geocodificados, ou permitir que o operador possa gerar as coordenadas x e y para o registro, simplesmente apontando a sua localização na tela do computador.

2.2.4 – Os erros na geocodificação de endereços

A chave para o sucesso na geocodificação de endereços está em quão bem o operador e o *software* respondem em conjunto com os endereços não geocodificados. Com um banco de dados de referência bem construído e um banco de dados de endereços alvo livre de erros, o processo de geocodificação é trivial. Entretanto, no mundo real existem erros em ambos os bancos de dados, o que significa que um percentual de correspondência de 100% só é possível em bancos de dados muito pequenos.³⁸

Existe um número significativo de erros potenciais presentes na geocodificação utilizando um banco de dados referencial de trechos de logradouros. Esses erros são provocados por diversas fontes de problemas, tanto no banco de dados de referência como no arquivo de endereços alvo. As principais fontes de problemas são listadas abaixo:

- **Desatualização do banco de dados de referência.** Os bancos de dados de referência podem estar desatualizados com a realidade de campo. Na verdade, a cidade é um espaço em constante mudança, principalmente em áreas de crescimento urbano acelerado. Isto pode significar que novos endereços em inúmeros novos logradouros são desconhecidos. Também é bastante comum, principalmente na realidade brasileira, a constante alteração da denominação do logradouro pelas câmaras municipais. Tais alterações podem originar confusão e incompatibilidade entre os bancos de dados de referência e o arquivo alvo. Também pode ocorrer a ausência de faixas de numeração ou faixas de numeração incorretamente atribuídas.

³⁸ DRUMMOND, W. J., "Address Matching", *Journal of the American Planning Association*; Sprint95, Vol. 61 Issue 2, p340, 11p.

Muitos endereços não são geocodificados por razão de haver mais de um trecho de logradouro com a faixas de numeração sobrepostas, ou mesmo sem a faixa de numeração.

- **Abreviações e erros gramaticais.** Os nomes de logradouros podem ser gravados incorretamente ou estar abreviados. Por exemplo: *Rua Doutor Ferreira de Carvalho* pode ser gravada como *Rua Dr Ferreira de Carvalho* ou como *R Dr. Ferreira de Carvalho*. Muitos *softwares* podem ser capazes de reconhecer somente algumas variações. Frequentemente, uma tabela de correspondência de abreviações e erros gramaticais necessita ser elaborada manualmente pelo operador.
- **Variações locais de nomes de logradouros.** Em algumas áreas o nome do logradouro no banco de dados de referência pode não refletir o nome do logradouro conhecido localmente (alguns logradouros são conhecidos por seus apelidos, como por exemplo: Av Catalão ao invés de Av Carlos Luz, podendo resultar na não geocodificação dos endereços.
- **Logradouros homônimos.** Em muitas localidades é comum encontrarmos o mesmo nome de logradouro em um ou mais bairros. Por exemplo, 30 ocorrências da Rua A podem ser encontradas no município de Belo Horizonte. Se o processo de geocodificação não especificar um bairro ou um outro atributo de limitação, é possível que um endereço não seja geocodificado por encontrar múltiplas possibilidades de localização.
- **Erros tipográficos.** Simples erros de tipografia podem mudar o endereço Rua Januária 30 em Rua Januária 300, um endereço que pode não existir.
- **Numeração irregular de imóveis.** Alguns logradouros possuem uma realidade que não obedece ao padrão de numeração. Particularmente comum no Brasil, a numeração dos imóveis pode não crescer uniformemente, bem como possuir em uma mesma face números pares e ímpares caoticamente distribuídos. Os endereços em logradouros que possuam esta realidade não podem ser geocodificados, uma vez que o modelo de geocodificação não prevê esta situação.
- **Sentido incorreto.** Um banco de dados de referência pode conter trechos de logradouros cujo sentido esteja incorreto. De acordo com o padrão de disposição da numeração ao longo do logradouro, os números pares e ímpares do logradouro deverão estar posicionados à direita e à esquerda do trecho de logradouro respectivamente, de acordo com o sentido do mesmo. Se este sentido não corresponder à realidade de campo, os endereços poderão não ser geocodificados, ou

ser geocodificados no lado errado do logradouro.

- **Imprecisão na geocodificação.** O ponto geocodificado pode estar a alguma distância de sua real localização. Esta imprecisão pode não ser muito relevante, dependendo da aplicação, mas é sempre desejável que o ponto não se situe muito distante da sua real localização e que não esteja, por exemplo, localizado fora dos limites da sua quadra. É importante enfatizar que o processo de interpolação, por natureza, pode produzir somente uma aproximação da localização para cada endereço. Um endereço qualquer pode ser interpolado para ocorrer a 1/3 ao longo do trecho de logradouro, mas sua verdadeira posição se localizar no meio do trecho, bastando que o endereço anterior ocupe um lote bastante largo.

Uma última consideração diz respeito à limitação espacial da geocodificação. Qualquer localização obtida pelo processo de geocodificação será aproximada e necessariamente limitada à exatidão do banco de dados de referência. Um banco de dados de referência, por exemplo, que tenha sido gerado pela digitalização de mapas na escala 1:100.000, poderá possuir pontos localizados aproximadamente a 50 metros de suas verdadeiras localizações na Terra (DRUMMOND, 1995).³⁹ Considerando as incertezas adicionais introduzidas pela interpolação, fatores de afastamento/encurtamento, entre outras, um endereço geocodificado em uma base de referência com este nível de qualidade não poderia ser utilizado quando é desejado um alto nível de exatidão espacial.

³⁹ DRUMMOND, W. J., "Address Matching", Journal of the American Planning Association; Sprint95, Vol. 61 Issue 2, p340, 11p.

2.2.5 – Tratamento dos registros não geocodificados

Após uma operação de geocodificação, os registros de endereços do arquivo alvo podem ser teoricamente divididos em quatro classes (DRUMMOND, 1995).⁴⁰ A primeira classe abrange as correspondências perfeitas, ou seja, aqueles registros que contém as coordenadas x e y adequadas (aproximadas) para o endereço alvo. A segunda classe contém os registros não correspondidos que não poderiam ser geocodificados em nenhuma circunstância, o que poderia ser verdade para um registro cujo endereço está faltando, mutilado, etc. Tal situação também poderia resultar para os registros cujos endereços se encontram em uma área recentemente construída e que não está incluída no banco de dados de referência. A terceira e quarta classes descrevem dois tipos de erros fundamentais: a correspondência incorreta (falso positivo), que inclui os registros que foram correspondidos, por alguma razão, ao trecho de logradouro incorreto, e a não correspondência (falso negativo) devida a algum pequeno erro, como a omissão do bairro.

Na maioria dos casos, de 25 a 75 por cento dos registros de endereços no arquivo alvo serão geocodificados com uma correspondência perfeita (DRUMMOND, 1995). O tratamento dos registros não geocodificados pode utilizar três estratégias:

1. Usar somente os registros com uma correspondência perfeita para análise e mapeamento e ignorar os registros que não foram perfeitamente correspondidos;
2. Aumentar o percentual dos registros geocodificados incluindo, além de todos os registros perfeitamente correspondidos, aqueles de correspondência parcial mais confiável;
3. Aumentar o percentual dos registros geocodificados incluindo, além de todos os registros perfeitamente correspondidos, aqueles de correspondência parcial confiável ou não.

A taxa de correspondência é o número total de correspondência (corretas e incorretas) dividido pelo número total de registros, enquanto a taxa de erro é o número de correspondências incorretas dividido pelo número total de correspondência. Assim, mover-se

⁴⁰ DRUMMOND, W. J., "Address Matching", Journal of the American Planning Association; Sprint95, Vol. 61 Issue 2, p340, 11p.

da estratégia 1 para a estratégia 2 para a estratégia 3, significa aumentar a taxa de correspondência por um lado, mas também, por outro lado, aumentar a taxa de erro.

Infelizmente, não há uma regra absoluta para determinar a melhor escolha entre as três estratégias acima. Para fazê-la, o analista deve avaliar os benefícios de uma taxa de correspondência mais alta, comparada com os custos de uma taxa de erro também mais elevada. Esta avaliação está intimamente ligada ao problema que se busca resolver. Por exemplo, no caso de despacho de serviços de emergência, as conseqüências de uma alta taxa de erro (veículos de emergência enviados para o local errado) são mais severas do que aquelas de uma taxa de erro menor, características da estratégia 1.

2.2.6 – Ferramentas para aumentar a taxa de correspondência

Existem alguns mecanismos que podem ser utilizados para produzir melhores taxas de correspondência no processo de geocodificação. Estes mecanismos podem ser colocados em duas categorias: ferramentas de transformação e estratégias de atuação. As ferramentas de transformação são aplicadas ao banco de dados alvo e/ou ao banco de dados de referência para converter a informação de endereço em uma forma mais provável de se obter uma correspondência. Pode-se, por exemplo, reordenar os registros alvo, transformar nomes para letras maiúsculas, retirar vírgulas e acentos. Também se utilizam tabelas de conversão da informação de tipo de logradouro, sinônimos de nome de logradouros e abreviações. Uma tabela de transformação poderia converter o tipo de logradouro “Rua”, “R.”, “RU” para uma forma padrão “R”, por exemplo.

Uma outra poderosa ferramenta de transformação, é a chamada função *soundex* (Knuth, 1973).⁴¹ Esta função cria um equivalente fonético para a ortografia de cada nome de logradouro, permitindo que o *software* de geocodificação compare os nomes baseado nos valores equivalentes, ao invés de sua ortografia literal. Por exemplo, os nomes de logradouros “Flamboyant”, “Flamboiant” e “Flamboiam” poderiam ter o mesmo equivalente fonético “FLMBN”. O uso desta ferramenta pode aumentar substancialmente a taxa de correspondência, mas também pode fazer crescer a taxa de erro a níveis intoleráveis. Sempre que se utilizar uma ferramenta como esta, deve-se usar atributos adicionais (como o CEP) e examinar cuidadosamente o resultado de uma amostra produzida para medir sua

⁴¹ KNUTH, D. E., “The Art of Computer Programming”, 1973, Volume 3: Sorting and Searching. Reading, MA: Addison-Wesley.

confiabilidade.

Os *softwares* de geocodificação podem aumentar a taxa de correspondência através de estratégias de atuação, ou seja, critério de relaxamento, tabelas de pontuação e análises probabilísticas. O critério de relaxamento requer múltiplas execuções do processo de geocodificação com sucessivos relaxamentos dos requisitos de correspondência. A primeira execução pode requerer uma correspondência exata do nome do logradouro, número do imóvel, tipo de logradouro, nome do bairro e código postal. Uma segunda execução pode ignorar o nome do bairro. Uma última execução pode substituir a função *soundex* pelo nome do logradouro. Durante cada execução o operador especifica quais critérios são aplicados, e a aplicação de todos os critérios determina se a correspondência teve ou não sucesso.

As tabelas de pontuação aplicam, em primeiro lugar, o algoritmo *soundex* ao nome de logradouro para, então, gerar uma lista de trechos de logradouros candidatos. Cada candidato inicia com uma pontuação de 100, mas o *software* de geocodificação deduz um número de pontos especificado pelo operador para cada elemento não correspondido. A ausência de um bairro ou um nome de bairro diferente poderia ter uma penalidade de 1 ponto; uma correspondência exata do nome de logradouro poderia não ter penalidade, mas uma correspondência através da função *soundex* poderia ter uma penalidade de 3 pontos. O trecho de logradouro com a maior pontuação poderia, portanto, ser aquele escolhido para a geocodificação do endereço, desde que este tenha alcançado uma pontuação mínima especificada pelo operador.

A abordagem probabilística cria correspondências baseadas em uma sofisticada teoria de correspondência de registros (JARO, 1989).⁴² De forma resumida, o banco de dados de endereço alvo e o banco de dados de referência são divididos em subconjuntos baseados nos valores de um atributo comum como o CEP. Dentro de cada subconjunto cada registro alvo é comparado com cada registro de referência. Durante esta comparação do tipo “um-para-um”, são assinalados pesos individuais para cada atributo de relacionamento do registro alvo (número do imóvel, bairro, etc.) baseados nas similaridades com o registro de referência.

⁴² JARO, Matthew A, “Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida”, *Journal of the American Statistical Association*, 84,4:414-20,1989.

Os pesos são somados e produzem uma pontuação total representando a probabilidade que um dado registro alvo possui de estar localizado em um determinado trecho de logradouro. As comparações cuja pontuação total se enquadra acima de um determinado valor de referência determinado pelo operador são consideradas correspondência positiva; aquelas que se enquadram abaixo de um segundo valor, correspondência negativa; aquelas que se enquadram entre os dois valores de referência devem ser revistas interativamente pelo operador. As correspondências negativas são novamente divididas em subconjuntos por um atributo diferente (tal como o valor para a função *soundex* para o nome do logradouro), e o processo é novamente repetido.

As três estratégias de atuação acima descritas ilustram uma situação clássica de negociação entre a facilidade de uso e poder. O critério de relaxamento é o método mais fácil para entender e usar. A abordagem estatístico-probabilística é o método mais poderoso, flexível e estatisticamente válido, embora seja também o mais difícil a ser entendido e utilizado. O método de tabelas de pontuação combina boa facilidade de uso com uma moderada flexibilidade. A escolha de qualquer um desses métodos irá depender, mais uma vez, das necessidades de geocodificação e das características do problema que se quer resolver.

2.3 – Metodologia e Técnicas Utilizadas

Este trabalho utilizou uma metodologia específica de desenvolvimento de aplicativos computacionais, que incorporou as técnicas inerentes aos Sistemas de Informações Geográficas e a Metodologia de Leitura de Consumo e Entrega de Faturas adotada na COPASA. O desenvolvimento do aplicativo no ambiente GIS utilizou os conceitos e procedimentos que formam a análise essencial de sistemas ⁴³ de McMenamin e Palmer ⁴⁴ e consistiu das fases de especificação e projeto físico. Para a construção do modelo essencial do sistema foram utilizadas algumas ferramentas da análise estruturada (digrama de fluxo de dados, dicionário de dados e português estruturado) e da modelagem da informação (diagrama entidade-relacionamento).

2.3.1 – Especificação

A Especificação conceitua o sistema do ponto de vista lógico. Esta é a fase de maior abstração, onde pouca ou nenhuma consideração é feita acerca da implementação do sistema. Esta fase consiste das etapas abaixo descritas:

- **Levantamento de dados**

A primeira etapa na especificação consiste da obtenção de conhecimento, onde se busca, em última análise, um entendimento da realidade acerca da necessidade ou do problema a ser solucionado. Existem várias técnicas utilizadas para a obtenção deste conhecimento, entre elas, as mais comuns são: técnicas de entrevista, levantamento de campo, *brainstorm* e visitas técnicas.

Para este trabalho esta etapa buscou basicamente o conhecimento através de:

- Levantamento dos processos executados no Sistema Comercial (SICOM) ⁴⁵ da

⁴³ A Análise Essencial de Sistemas é um método que permite descobrir e definir os requerimentos verdadeiros de um sistema. O método está baseado em um arcabouço conceitual composto por três elementos: uma descrição das características dos sistemas, uma classificação dos componentes da essência de um sistema, e uma classificação das restrições de implementação.

⁴⁴ MCMENAMIN, Stephen; PALMER, John, "Análise Essencial de Sistemas"; tradutor Lars Gustav Erik Unonius; revisor técnico Fernando Manso, São Paulo, Ed. McGraw-Hill, 1991.

⁴⁵ O Sistema Comercial (SICOM) é responsável pela manutenção do cadastro de clientes, pelo faturamento, arrecadação e

COPASA para a manutenção do seu cadastro de clientes;

- Visitas técnicas ao setor de cadastro comercial e faturamento da COPASA, para entendimento dos processos executados;
- Levantamento de campo para entendimento da realidade;

Uma segunda etapa consiste de uma revisão da bibliografia existente sobre o assunto envolvido. Nesta etapa, foram revisadas as bibliografias sobre Sistemas de Informações Geográficas, estruturação de bancos de dados, métodos de desenvolvimento de sistemas de informação, a geocodificação de endereços e interpolação.

• **Modelagem do Problema**

A modelagem do problema visa a representação do problema da geocodificação de endereços com uma abordagem que utiliza a metodologia de leitura de consumo e entrega de faturas da COPASA, aliada aos conceitos de *GIS*.

• **Elaboração do Modelo Ambiental**

O Modelo Ambiental visa estabelecer uma visão geral do sistema, definindo seus objetivos, suas fronteiras e sua interação com o ambiente. É o conjunto balanceado dos seguintes produtos:

- Declaração dos objetivos do sistema;
- Diagrama do Contexto;
- Lista de Eventos;

• **Elaboração do Modelo Comportamental**

A elaboração do modelo comportamental visa definir o sistema do ponto de vista interno, descrevendo como o sistema reage internamente aos estímulos do exterior, enquanto um conjunto de elementos inter-relacionados, como um todo organizado. Consta dos seguintes produtos:

- Elaboração do diagrama de fluxo de dados (DFD);
- Elaboração do diagrama de entidade-relacionamento (DER);

- Elaboração das mini-especificações e diagramas de transição de estados.

2.3.2 – Projeto físico

O projeto físico é responsável pela estruturação física do sistema onde todas as especificações da fase anterior são implementadas. É nesta fase que os requisitos do sistema são efetivamente construídos e o sistema de informações passa a ser operado. As etapas abaixo descritas resumem as principais atividades desta fase:

- **Arquitetura física de implementação**

O projeto da arquitetura física de implementação é uma das etapas do projeto físico e trata da identificação dos recursos de hardware, software básico, recursos humanos e logísticos que serão utilizados pelo sistema. Nesta etapa é feito um modelo de implantação envolvendo todos os recursos necessários à operação do sistema.

- **Projeto físico das bases de dados**

O projeto físico das bases de dados é a etapa que trata da elaboração do modelo físico das bases de dados baseado nas necessidades do sistema e nas funcionalidades do ambiente de *software* e *hardware* a ser utilizado.⁴⁶ Compõe-se dos seguintes procedimentos:

- Derivação do modelo conceitual de dados para o modelo de dados característico do ambiente de *software* e *hardware* utilizado;
- Descrição dos *layouts* dos registros, com identificação das chaves primárias, domínios dos atributos e chaves de acesso necessárias;
- Incorporação dos dados de controle necessários ao funcionamento do sistema;
- Definição dos critérios de segurança e integridade das bases de dados.

de seus clientes e demais informações produzidas.

⁴⁶ O ambiente de software utilizado neste trabalho é o ARC/INFO Workstation 8.3 na plataforma Unix AIX 4.3.

- **Estruturação física do sistema**

A estruturação física do sistema é outra etapa que visa detalhar a estrutura física dos programas do sistema, estabelecendo a seqüência de execução dos processos definidos. Nesta etapa é elaborado o fluxograma dos módulos, discriminando os programas e/ou procedimentos, entradas e saídas estruturadas para o sistema.

- **Especificação das entradas e saídas**

A especificação das entradas e saídas é a etapa que visa definir o conteúdo e elaborar os *layouts* dos documentos, telas e relatórios (incluindo mapas temáticos) do sistema.

- **Definição dos programas (algoritmos)**

A definição de programas ou algoritmos é a etapa que visa definir os programas que implementarão os processos do sistema, incluindo a elaboração do código fonte (algoritmos).

- **Elaboração e teste de programas**

A elaboração e teste de programas é a etapa que visa codificar e testar individualmente os programas do sistema.

- **Teste integrado**

O teste integrado do sistema é a etapa que visa garantir a correta execução dos módulos dentro do contexto geral do sistema. É composta dos seguintes procedimentos:

- Realização dos testes utilizando a documentação do sistema;
- Balizamento dos resultados com a realidade;
- Realização dos ajustes necessários.

Capítulo III – Desenvolvimento do Trabalho

3.1 - Introdução

3.1.1 – Descrição da realidade

O município de Belo Horizonte possui um total de 11.894 logradouros oficiais ⁴⁷ com aproximadamente 540.000 imóveis registrados no cadastro de clientes da COPASA ⁴⁸, entre consumidores residenciais, comerciais, industriais e públicos.

Origem do cadastro de Clientes ⁴⁹

O sistema de saneamento é fundamentalmente um sistema de distribuição em grande escala sobre uma base geográfica. Apesar de a água ser um produto armazenável, o sistema deve necessariamente levar em conta duas variáveis: a variável geográfica, da distribuição territorial dos consumidores e a variável temporal, representada pela demanda específica de cada consumidor, que varia ciclicamente durante o dia, a semana e o ano, além de poder apresentar uma variação de tendência, em geral crescente, ao longo dos anos.

Para documentar as variáveis geográfica e temporal, essenciais para o planejamento das redes de distribuição e para o faturamento, os cadastros nasceram como livros contábeis, nos quais eram agrupados os consumidores. Cada livro ou razão representava uma área geográfica que agrupava um número variável de rotas ou roteiros. O leitorista possuía um livrete, no qual anotava o consumo, e cada livrete agrupava o número de consumidores que o leitorista conseguia ler em um dia de trabalho. Assim, o registro mensal do consumo de água de cada consumidor, os livros, os roteiros e os livretes, representavam, grosso modo, a distribuição do consumo em base geográfica e sua evolução no tempo.

As estruturas urbanas crescem freqüentemente sem levar em conta a distribuição dos serviços essenciais, tornando instável a base geográfica sobre a qual é montado o cadastro. A água é um dos primeiros serviços públicos fornecidos em bairros novos e isso faz com que a ligação possa ser efetuada antes do arruamento oficial e, portanto, da numeração das casas.

⁴⁷ Fonte: Prefeitura de Belo Horizonte, 2004.

⁴⁸ Fonte: COPASA, 2005.

⁴⁹ FUSP: "Metodologia de Análise de Cadastros de Concessionárias", São Paulo, Convênio ANEEL/FUSP, 2000.

A consequência disto, em algumas áreas de distribuição, é que se pode encontrar até três numerações e três nomes de ruas diferentes para o mesmo endereço físico: endereço dos serviços elétricos, endereço dos serviços da água e endereço oficial da prefeitura. Também é encontrada a situação em que os logradouros possuem uma numeração irregular, ou seja, uma numeração que não segue um padrão, o que dificulta extremamente a localização do endereço. Esta situação, para garantir a leitura mensal regular dos consumos, obriga o concessionário a utilizar o roteiro como verdadeiro guia geográfico, no qual o código de leitura é um número seqüencial que representa a ordem que o leiturista segue para a leitura. Às vezes, o roteiro é até representado como uma seqüência de códigos colocados sobre um mapa, que representa o percurso físico do leiturista no processo de controle mensal.

O cadastro de clientes da COPASA

O cadastro dos clientes comerciais é mantido pelas áreas comerciais regionalmente distribuídas de acordo com a divisão administrativa da COPASA. Para a localidade de Belo Horizonte, contexto de implantação do SIURB, o setor de controle de cadastro da Divisão de Faturamento é a área responsável pela manutenção do cadastro de clientes, inclusive pela consistência das informações de endereço, tais como: tipo e nome de logradouro, bairro, CEP e numeração de porta. Já a Divisão de Operações Comerciais da Metropolitana – DVOC é a área responsável pela distribuição destes clientes em rotas de leitura.

Até o desenvolvimento do sistema SIURB os clientes da COPASA não possuíam as coordenadas geográficas como identificador de localização, somente o endereço descritivo. Este endereço é obtido com a solicitação de ligação de água ou esgoto pelo cliente que informa o endereço para a ligação. A primeira necessidade de informação deste processo é o conhecimento da viabilidade da ligação, ou seja, existe uma rede suficientemente próxima ao endereço informado? Esta informação é obtida basicamente de três formas:

- Solicitação de informação ao cliente baseado na pergunta: Existe uma ligação de água/esgoto imediatamente vizinha a este endereço? Neste caso, confia-se na informação prestada pelo cliente, que pode não ser correta ou precisa;
- Consulta ao cadastro de clientes existentes cujo número de porta é próximo ao endereço informado;
- Visita em campo ao endereço informado para obtenção ou confirmação da existência

de viabilidade econômico-operacional da ligação.⁵⁰

Uma vez constatada a viabilidade da execução da ligação de água/esgoto, a ordem de serviço de ligação de água/esgoto é emitida e o serviço de ligação executado. Após a efetivação da ligação o cliente é incluído no sistema comercial – SICOM.⁵¹ Além das informações cadastrais pertinentes, duas informações essenciais ao processo comercial são acrescentadas: a matrícula e o *localizador*.⁵²

A matrícula é um atributo numérico que identifica o cliente univocamente em todos os processos comerciais e não sofre alteração. Como é um atributo imutável e não existe a possibilidade de duplicidade, a matrícula é utilizada como chave primária no relacionamento com outros arquivos.

O *localizador* é um atributo numérico de localização do imóvel que possui uma ligação de água e/ou esgoto. Diferentemente da matrícula o *localizador* pode sofrer alteração, apesar de identificar univocamente o imóvel. O processo de alteração do *localizador* é conhecido como remanejamento.

O remanejamento de localizador

O *localizador* pode ser alterado devido a inúmeros fatores, entre eles:

- Excesso de imóveis em uma rota de percurso do leiturista (rota excessivamente grande para uma jornada diária de leitura e de entrega de faturas);
- Inclusão de novos clientes por crescimento vegetativo ou novos loteamentos;
- Redimensionamento de rotas promovido por novas políticas comerciais ou devido à implantação de novas tecnologias de faturamento (emissão de faturas *online*, por exemplo);

⁵⁰ Esta viabilidade diz respeito somente à existência de um trecho de rede de água ou de esgoto suficientemente próximo ao endereço da ligação de forma a viabilizar economicamente a ligação. Caso haja a necessidade de um prolongamento de rede, este é negociado com o cliente. Na ligação de água, a viabilidade concernente à existência de pressão e vazão suficientes é normalmente realizada em uma etapa anterior no processo de implantação ou expansão de trechos de rede.

⁵¹ Os processos de ligação de água e/ou esgoto e de inclusão do cliente no SICOM não informam as coordenadas geográficas do imóvel.

⁵² Ver item 3.1.3: A metodologia de leitura de consumo e entrega de faturas da COPASA para uma descrição do localizador.

Independentemente do fator que provoca a alteração do *localizador*, existe uma grande dificuldade no dimensionamento e na melhor distribuição das rotas de leitura. Uma das principais razões desta dificuldade está na inexistência de uma visão espacial das rotas, principalmente quando os remanejamentos são necessários em uma grande área geográfica. Este é um processo que normalmente pode levar dias e está sujeito a erros que só são corrigidos após a implantação das alterações. Para promover estas alterações a DVOC dispõe do tempo médio de leitura e entrega de fatura por imóvel, além dos mapas de rotas que são croquis referentes a uma rota de leitura. Frequentemente estes mapas são apenas esboços aproximados da forma e do traçado da realidade do percurso em campo, tornando-se fontes pouco confiáveis de informações sobre as rotas existentes.

O cadastro de clientes da COPASA consiste de uma tabela contendo todas as informações comerciais do cliente real, factível ou potencial da empresa nas localidades onde ela atua. O cliente real é aquele que possui uma ligação de água e/ou esgoto. Os clientes do tipo factível ou potencial são aqueles que não possuem uma ligação de água e/ou esgoto. No caso do cliente factível, existe uma viabilidade técnica de se efetuar de imediato uma ligação de água e/ou de esgoto que não existe para o cliente potencial. Esta tabela de clientes é mantida pelo sistema comercial implantado no ambiente de computador de grande porte da empresa (*mainframe*) e gerenciado pelo software gerenciador de banco de dados ADABAS.⁵³ O cadastramento dos clientes da COPASA se inicia com o pedido de ligação de água e/ou esgoto.

O pedido de ligação de água e esgoto

Ao solicitar uma ligação, o cliente informa o endereço do imóvel e o setor de atendimento e serviços da COPASA abre uma ordem de serviço de ligação para o distrito de serviço responsável pelo atendimento da região. O distrito de serviço, após a verificação da viabilidade técnica, realiza o serviço de ligação e retorna para o setor comercial incluir o cliente no sistema. O setor comercial, por sua vez, determina o *localizador* do imóvel e inclui o cliente no sistema comercial - SICOM.

⁵³ O software ADABAS é um gerenciador de banco de dados, marca registrada da empresa Software AG.

Uma vez incluído o cliente, ele passa a ser visitado pelo leiturista no processo de leitura de consumo e entrega de faturas. Sempre que neste processo o número de imóveis ultrapassa o limite máximo desejável, tornando a rota muito extensa para o leiturista, ocorre o chamado remanejamento de setor-rota, onde ocorre uma modificação das rotas existentes e/ou a criação de novas rotas.

Assim, o cadastramento do cliente atende exclusivamente às rotinas do processo comercial de faturamento da COPASA. As informações espaciais ficam restritas ao endereço do imóvel e ao *localizador*. Ambas, juntamente com as demais informações comerciais do cliente, estão armazenadas na forma de uma tabela alfanumérica no banco de dados comercial. Para a inclusão e consistência do endereço são mantidas atualizadas pelo setor de cadastro comercial as tabelas de logradouros e bairros. Existe, entretanto, uma outra base de dados que contém informações a respeito dos logradouros e bairros, que é a base de dados geográfica utilizada pelos sistemas de informações geográficas da empresa. Nesta base, encontramos os *layers* de trechos de logradouros e bairros.

O layer trecho de logradouros

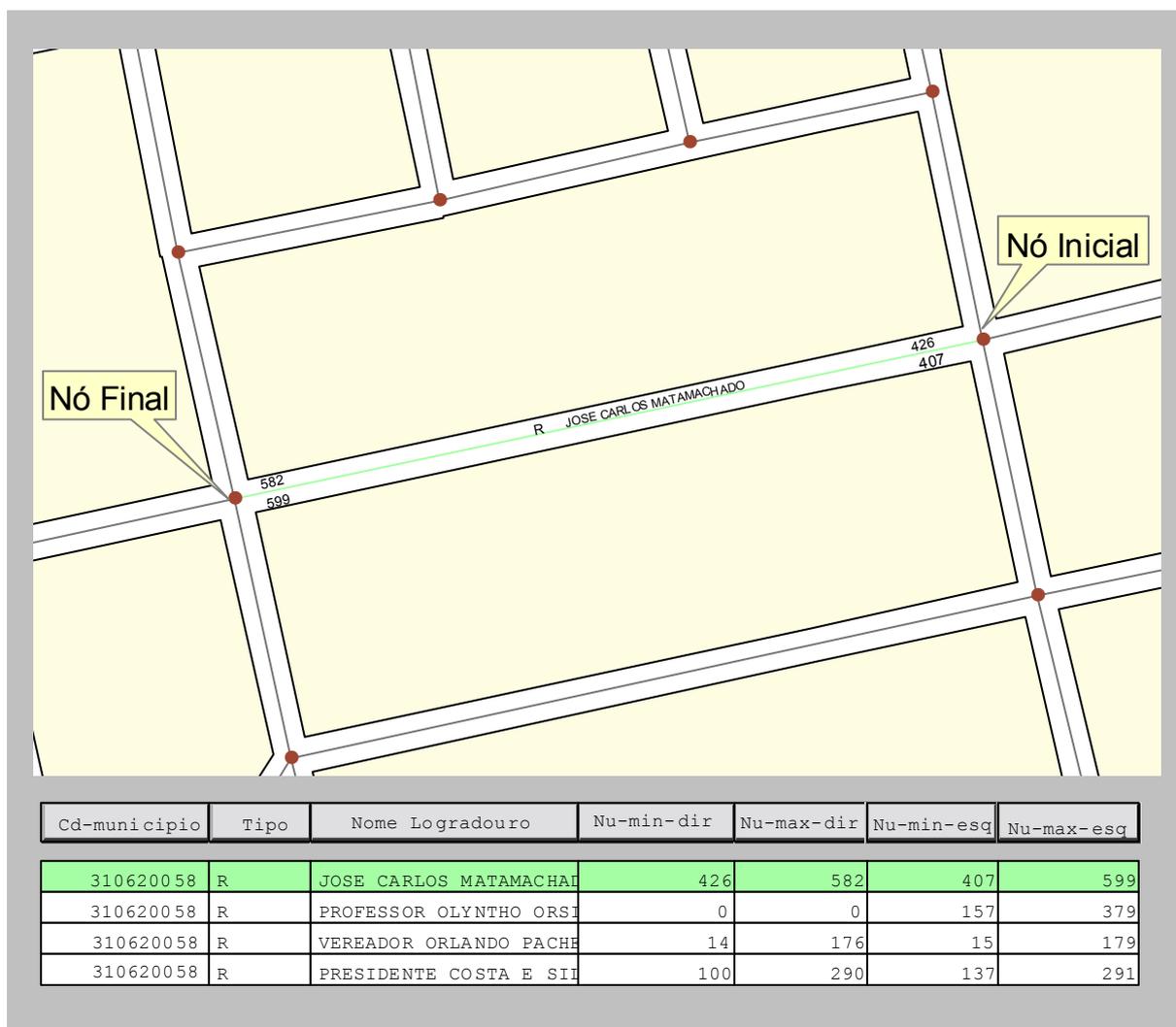
A COPASA possui a base de trechos de logradouros oficial da prefeitura de Belo Horizonte em meio digital, atualmente utilizada para servir de referência para o lançamento das redes de água e esgoto através do sistema de informações geográficas denominado SIGRAE⁵⁴. Este cadastro é composto pelas linhas imaginárias do eixo do logradouro subdividido em cada cruzamento. Um logradouro é assim representado pelos trechos que o compõe. Cada trecho possui os seguintes atributos básicos:

CD-MUNICÍPIO	Código do município a que pertence
CD-LOGRADOURO	Código do logradouro adotado pela COPASA
NO-LOGRADOURO	Nome do logradouro
TP-LOGRADOURO	Tipo do logradouro (Rua, Avenida, Beco, etc)
NU-MIN-DIR	Numeração mínima à direita do trecho
NU-MAX-DIR	Numeração máxima à direita do trecho
NU-MIN-ESQ	Numeração mínima à esquerda do trecho
NU-MAX-ESQ	Numeração máxima à esquerda do trecho

Tabela 3.1.1 – Atributos básicos dos trechos de logradouros.

⁵⁴ Sistema de Informações Geográficas das Redes de Água e Esgoto da COPASA. Este sistema é responsável pela atualização do cadastro de redes de água e de esgoto.

Os trechos de logradouros possuem uma estrutura derivada dos arquivos TIGER⁵⁵, ou seja, são uma coleção de segmentos de linhas nas quais se associam o nome do logradouro e uma faixa de numeração de imóveis em cada lado do logradouro como atributos. A figura 3.1.1 apresenta um exemplo da base de trechos de logradouros existente onde uma linha para o logradouro José Carlos Matamachado possui um nó de início, um nó de término e outros atributos. Os atributos para a linha indicam que números ímpares estão do lado esquerdo do logradouro e variam de 407 a 599.



⁵⁵ Ver item 2.2.2 – “Introdução” para maiores detalhes.

Figura 3.1.1 – Exemplo de eixo de logradouro com seus atributos associados.

Muitos fabricantes de softwares GIS construíram ferramentas derivadas dos arquivos TIGER para interpolar uma coordenada x e y de um determinado endereço. O endereço é utilizado para identificar um segmento de linha que tenha uma correspondência exata do nome do logradouro e cujo número do imóvel esteja contido nos limites de numeração do trecho.

A COPASA utiliza a ferramenta de *geocoding* do software ARC/INFO em todos os seus sistemas de informações geográficas para a localização de um determinado endereço dentro de um sistema de coordenadas.

Os Sistemas de Informações Geográficas da COPASA

O primeiro sistema de informações geográficas da COPASA foi implantado em 1996 através do desenvolvimento do Sistema de Informações Geográficas das Redes de Água e Esgoto – SIGRAE – que possibilita a manutenção de um cadastro com aproximadamente 5.000 km de redes de água e 3.000 km de redes de esgoto sanitário.

O SIGRAE possui 34 níveis de informação descritos na Tabela 3.1.2 a seguir:

Nome	Descrição	Tipo de feição
Municípios	Contorno do município de Belo Horizonte.	Polígono
Regionais	Contorno das regionais administrativas da Prefeitura de Belo Horizonte.	Polígono
Distritos	Contorno dos distritos operacionais da COPASA MG em Belo Horizonte.	Polígono
Bacias	Contorno das bacias de esgoto, segundo divisão da COPASA MG em Belo Horizonte.	Polígono
Sub-bacias	Contorno das sub-bacias de esgoto, segundo divisão da COPASA MG em Belo Horizonte.	Polígono
Zonas-Abast	Contorno das zonas de abastecimento, segundo divisão da COPASA MG em Belo Horizonte.	Polígono
Massas-agua	Contorno dos lagos e represas de Belo Horizonte.	Polígono
Cursos-dagua	Contém a hidrografia existente no município de Belo Horizonte.	Linha
Vias	Contorno dos meios-fios ou da delimitação existente em um arruamento do município de Belo Horizonte.	Linha
Quadras	Linha imaginária identificadora da quadra (quarteirão) do município de Belo Horizonte.	Polígono
Pinguelas	Contém as pequenas travessias em cursos d'água do município de Belo Horizonte.	Ponto
Bueiros	Contém os postos de visita das redes pluviais, de água e de esgoto de Belo Horizonte.	Ponto
Postes	Contém os postes da CEMIG em Belo Horizonte.	Ponto
Endereços	Correspondem aos imóveis da Prefeitura de Belo Horizonte.	Ponto

Nome	Descrição	Tipo de feição
Trechos	Linha imaginária (<i>Centerline</i>) identificadora do trecho de logradouro do município de Belo Horizonte.	Linha
Bairros	Contorno dos bairros de Belo Horizonte.	Polígono
Cercas e Muros	Contorno das cercas e muros dos imóveis de Belo Horizonte.	Linha
Edificações	Contorno dos imóveis de Belo Horizonte	Linha
Piscinas	Contorno das piscinas existentes em Belo Horizonte.	Linha
Cotas	Corresponde aos pontos cotados existentes na Prefeitura de Belo Horizonte.	Ponto
Curvas5	Curvas de nível de 5 em 5 metros de Belo Horizonte.	Linha
Raster 1:2000	Divisão das quadrículas na escala 1:2000.	Polígono
Raster 1:5000	Divisão das quadrículas na escala 1:5000.	Polígono
Redes CEMIG	Contém as redes subterrâneas da CEMIG.	Linha
Redes TELEMAR	Contém as redes subterrâneas da TELEMAR.	Linha
Redes GASMIG	Contém as redes da GASMIG.	Linha
Quadras URBEL	Contém as quadras de vilas e favelas da URBEL.	Polígono
Centerline URBEL	Contém as linhas imaginárias (<i>Centerline</i>) dos becos das vilas e favelas da URBEL.	Linha
Redes de Esgoto	Contém as redes de esgoto da COPASA MG em Belo Horizonte.	Linha
Caixas de manobra	Contém as caixas de manobra da COPASA MG em Belo Horizonte.	Polígono
Redes de água	Contém as redes de água da COPASA MG em Belo Horizonte.	Linha
Pontos-Notav	Contém os pontos que a COPASA MG considera especiais (notáveis) do município de Belo Horizonte. Ex.: hospitais, clínicas, escolas, etc.	Ponto
Esgcrit	Contém as áreas críticas de serviços de esgoto no município de Belo Horizonte.	Polígono
Aguacrit	Áreas que possuem abastecimento crítico de água em Belo Horizonte.	Polígono

Tabela 3.1.2 - Níveis que compõem a base de dados do SIGRAE.

Além do SIGRAE, foram posteriormente desenvolvidos os seguintes sistemas:

- Sistema CONECT - visa determinar os registros ótimos a serem fechados e as respectivas redes atingidas pelas manobras destes registros, no caso de interrupções e manutenções nas redes de água do município de Belo Horizonte;
- Sistema SIMAR - visa gerenciar as ordens de serviço de manutenção de redes de água e esgoto;
- Sistemas CONECTAGUA - visa prover informações quanto à conectividade da rede de água;
- Sistema CONECTESQ – visa prover informações quanto à conectividade da rede de esgoto.

Todos os sistemas trabalham no sistema de projeção UTM, datum South American 1969, fuso 23.

3.1.2 – O Problema da geocodificação de endereços

O método utilizado pela COPASA na geocodificação dos endereços é o método da interpolação das coordenadas geográficas dos endereços em uma base de dados de referência de trechos de logradouros.⁵⁶ Este método utiliza o endereço como parâmetro para a identificação do trecho de logradouro a ser considerado na interpolação. Todos os sistemas de informações geográficas implantados na COPASA utilizam este método para a localização de um endereço qualquer em um mapa. Um das principais questões a serem avaliadas em qualquer método de geocodificação de endereços é a **taxa de correspondência**. Entretanto, outras questões também devem ser avaliadas como a **imprecisão da geocodificação**, a **numeração irregular de logradouros** e o **acompanhamento das alterações cadastrais**.

A taxa de correspondência

Para a produção de uma base de dados geográfica dos clientes, como o proposto por esta dissertação, o volume de informações é muito grande e passa ser necessário executar uma geocodificação automática em lotes. Neste processo, as coordenadas são obtidas para os endereços nos quais houve correspondência exata de localização na base de dados de referência. Passa a ser necessário, então, lidar com os endereços rejeitados, ou seja, os endereços que não foram automaticamente geocodificados. Quando o processo for interativo, o trabalho com os rejeitados é manual. O endereço “ruim” é exibido juntamente com as comparações mais próximas presentes no banco de dados. O operador utiliza essas opções para selecionar a comparação mais provável. Isso envolve alguma subjetividade e a possibilidade de erros de geocodificação. Se o endereço introduzido for, por exemplo, Rua Cristina, 1234, e a única referência no banco de dados for Avenida Cristina, então é provável que relacionar o endereço ao trecho da Avenida Cristina seja o procedimento correto. Por outro lado, se o banco de dados contém Alameda Cristina, Avenida Cristina, entre outros tipos de logradouros, o relacionamento ao tipo “Avenida” pode estar errado. Este procedimento manual, entretanto, torna-se muito lento e oneroso quando o número de rejeitados for muito elevado. Se considerarmos uma taxa de correspondência de 80% na base de clientes da COPASA em Belo Horizonte, por exemplo, teríamos o equivalente a mais de

⁵⁶ Ver item 2.2.3: “O processo de geocodificação de endereços” para a descrição deste método.

100.000 endereços rejeitados. Para se tratar tamanho volume de registros, portanto, seria também necessário um tratamento automático em lote para os endereços não geocodificados. No item 2.2.5 são descritas algumas ferramentas para aumentar a taxa de correspondência, o que tende a minimizar o problema, embora aumente o risco de correspondências incorretas. Algumas estratégias podem aumentar a taxa de 55%, para 80% ou 95%, mas a taxa de correspondência errada também salta de 9% para 12% e 21%, respectivamente.⁵⁷ É um desafio, portanto, gerar taxas de correspondência elevadas sem que se elevem também as taxas de erro.

A imprecisão da geocodificação

Uma outra questão a ser avaliada é o problema da imprecisão da geocodificação. As coordenadas do endereço geocodificado devem estar, tanto quanto possível, dentro dos limites do lote ou, ao menos, dentro dos limites das quadras. Embora possa não ser muito relevante, dependendo da aplicação, é normalmente desconfortável observar um imóvel representado por um ponto localizado no meio da via pública, ou no lado errado da mesma. Deve-se buscar, portanto, mecanismos que venham a minimizar esses erros de posicionamento.

A numeração irregular de logradouros

Segundo dados da Prefeitura Municipal, cerca de 5% dos logradouros de Belo Horizonte possuem uma numeração irregular, ou seja, possuem uma numeração que não obedece ao padrão de numeração oficial, podendo ocorrer números aleatórios em ambas as faces do logradouro. Desta forma, não é possível modelar esta realidade no banco de dados de referência, que exige uma faixa de numeração regular em cada trecho de logradouro. Temos, assim, pelo método de geocodificação adotado pela COPASA, potencialmente uma taxa de rejeição mínima de 5% dos endereços.

⁵⁷ DRUMMOND, W. J., "Address Matching", Journal of the American Planning Association; Sprint95, Vol. 61 Issue 2, p340, 11p.

O acompanhamento das alterações cadastrais

Uma vez realizada, a geocodificação deve acompanhar as alterações cadastrais. Diariamente são realizadas inclusões, alterações e exclusões dos clientes no cadastro de clientes da COPASA. Muitas destas alterações são devidas à dinâmica natural das transformações urbanas, bem como ao processo interno de remanejamento de rotas.⁵⁸ O processo de alteração cadastral é realizado pelo sistema comercial cujas alterações devem refletir-se no cadastro geográfico gerado pelo processo de geocodificação. Este processo, portanto, deve ser contínuo e permanente para que o cadastro se mantenha sempre confiável e atualizado.

Um dos problemas que surgem com as alterações no ambiente urbano é a desatualização gradual da base de dados de referência. Quando um novo endereço é incluído, por exemplo, pode acontecer que a numeração do imóvel seja superior ou inferior à numeração máxima e mínima cadastrada no trecho de logradouro ao qual este imóvel está localizado. Se este trecho de logradouro não for atualizado com a nova faixa de numeração, este imóvel será rejeitado em uma futura geocodificação do endereço. Ao longo do tempo, a base de referência passa a perder confiabilidade, principalmente em áreas de grande crescimento urbano.

A solução adotada para a geocodificação de endereços buscou minimizar ou mesmo eliminar os problemas acima descritos. A metodologia proposta eliminou a possibilidade de erro na correspondência dos endereços contra a base de dados de referência, utilizou os fatores de afastamento e encurtamento para minimizar a imprecisão da geocodificação por interpolação no trecho de logradouro, eliminou a possibilidade de rejeição de endereços em logradouros com a ocorrência de numeração irregular valendo-se metodologia de leitura de consumo e entrega de faturas que a COPASA adota, e buscou, através da implantação de um sistema de informações geográficas corporativo, manter a confiabilidade do banco de dados de referência ao garantir a sua constante atualização.

⁵⁸ Ver item 3.1.3: “A metodologia de leitura de consumo e entrega de faturas da COPASA” para maiores detalhes.

3.1.3 – A metodologia de leitura de consumo e entrega de faturas da COPASA

A COPASA adota uma metodologia de leitura de consumo e entrega de faturas baseada na subdivisão do espaço geográfico urbano em rotas de visitação. Para que o processo de leitura possa acontecer, a COPASA adota uma metodologia apropriada para possibilitar a visitação de todos os clientes cadastrados, independentemente dos problemas de organização do espaço geográfico que são comuns nas áreas urbanas.

O Processo de Leitura

Na COPASA, o processo de leitura é realizado nas seguintes etapas:

1. Para a realização da leitura, o leiturista se utiliza de um formulário específico ou de um aparelho eletrônico chamado coletor de dados. No formulário ou no aparelho encontram-se as suas informações cadastrais e é organizado na forma de um roteiro de leitura.
2. O leiturista anota no formulário ou digita no coletor a leitura que encontra no seu medidor. Ela é comparada com as informações dos meses anteriores. Se a leitura digitada apresentar um valor muito alto ou muito baixo na comparação com os meses anteriores, o leiturista verifica o medidor e anota ou digita novamente.
3. Concluída a coleta das leituras, o leiturista retorna à concessionária onde serão processadas as informações necessárias ao cálculo de sua fatura de consumo de água e coleta de esgoto. Quando são utilizados coletores eletrônicos, a fatura pode também ser calculada e emitida no ato da leitura.
4. Emitida a fatura, ela é posteriormente entregue na residência do cliente pelo mesmo processo de roteamento. Também pode ser enviada aos correios para o serviço de entrega ou entregue no mesmo instante pelo leiturista, no caso do coletor eletrônico com emissão de fatura.

O Roteiro de Leitura

Conforme mencionado, o roteiro de leitura é o guia no qual o leiturista se orienta geograficamente para realizar a leitura do consumo e a entrega de faturas. Este roteiro visa solucionar muito dos problemas relacionados à localização dos endereços, principalmente os

provocados pela inconsistência da numeração e pela numeração irregular.

Basicamente, os roteiros são divisões do espaço geográfico no qual estão distribuídos os consumidores. Sua formulação segue algumas regras. A primeira refere-se à capacidade do leitorista em percorrer uma determinada rota, ou seja, a rota deve obedecer aos limites físicos de percurso do leitorista em um dia de trabalho. Neste caso, são levados em consideração os aspectos da distância a percorrer e do número de leituras a realizar. Também são avaliadas as dificuldades do percurso, não só a topografia, como também a dificuldade de acesso. A segunda trata da otimização, ou seja, da busca do melhor traçado ou aquele de menor gasto energético. Neste sentido, estuda-se um trajeto ou rota na qual haja o menor número possível de travessias de ruas, busca-se o melhor ponto de início da rota afim de melhor aproveitar a topografia do terreno, etc. Outras regras se relacionam com os aspectos econômico-financeiros, ciclos de faturamento, entre outros, que fogem ao escopo de discussão deste trabalho.

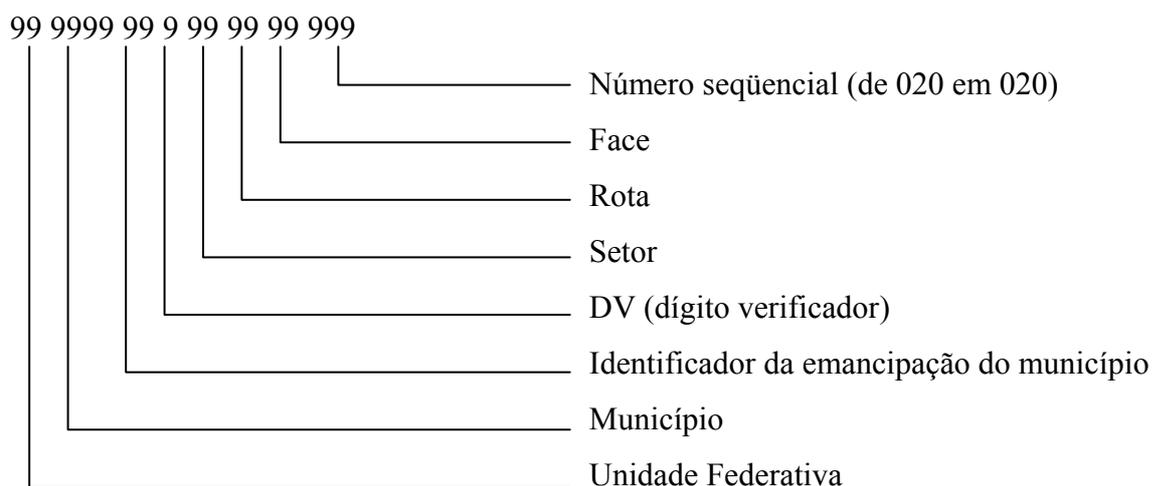
O processo de elaboração das rotas inicia-se na setorização da região que se deseja trabalhar. Ao setor definido atribui-se um código de duas posições numéricas que podem variar de 01 a 99. Assim, dentro de cada setor, traçam-se as rotas de acordo com as regras acima levantadas, e a cada uma também se atribui um código de duas posições numéricas que podem, igualmente, variar de 01 a 99.

A rota também é subdividida em faces, que normalmente correspondem aos quarteirões da quadra. Às faces também se atribui um código de duas posições numéricas que podem variar de 01 a 99. São nas faces que estão localizados os endereços dos imóveis, mas os mesmos não são identificados pela numeração, e sim, por um número seqüencial de três posições numéricas que podem variar de 001 a 999.

O leitorista percorre o trajeto de sua rota guiado pela estrutura seqüencial das faces e pelo conhecimento do ponto de partida. O imóvel é também identificado pela numeração seqüencial dentro da face, independente da numeração do imóvel (embora ele também possua esta informação). Elimina-se, assim, o problema da numeração irregular ou inconsistente, bastante comum nos centros urbanos. O roteiro de leitura é identificado por um código chamado de *localizador*.

Número do Localizador

O número do *localizador* é o conjunto de algarismos segundo o qual a COPASA identifica a localização dos imóveis cadastrados por ela. A tarefa de atribuir um *localizador* não segue ao acaso, pelo contrário, obedece a determinados critérios que visam facilitar a rápida localização daquele imóvel.



A figura 3.1.2 mostra a divisão em setores da região do distrito operacional sudoeste de Belo Horizonte. Se selecionarmos o setor 18 e conhecendo o número de identificação da localidade, que neste caso é 310620058, teremos o número 310620058-18. Com este número já temos a informação da Unidade Federativa, do município e setor. Mas não termina aí a identificação de um imóvel usuário. Observando outra vez o mapa da figura 3.1.2 vemos que cada setor abrange muitos logradouros. Para facilitar o trabalho do leiturista, estes grandes agrupamentos de logradouros, chamados setores, são divididos em agrupamentos menores chamados rotas. Assim, a rota é uma subdivisão de um setor. É caracterizada por um traçado que determina um roteiro a ser seguido pelo leiturista. Tal traçado permite que o leiturista, fazendo o percurso da rota do seu ponto inicial ao ponto final, visite no mesmo dia todos os imóveis daquela rota. Cada rota terá dois algarismos para ser identificada.

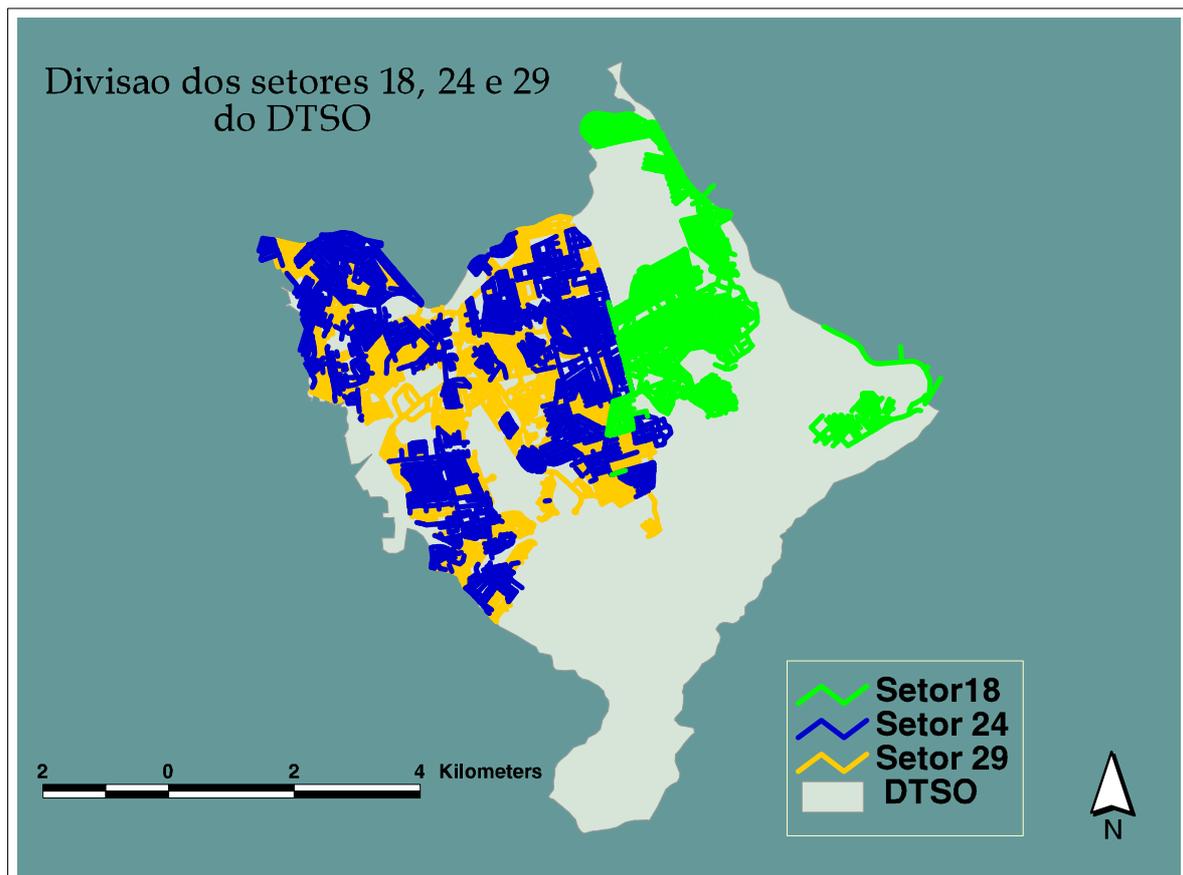


Figura 3.1.2 – Divisão em setores das rotas de leitura de consumo e entrega de faturas do distrito de serviços sudoeste da COPASA em Belo Horizonte.

A figura 3.1.3 mostra o mapa da rota 27 do setor 18. O número de rota, em conjunto aos outros já conhecidos, formará o número 310620058-1827 correspondente à localidade, setor e rota. Entretanto, ainda não teremos um *localizador* completo porque toda rota é dividida em faces. A face é a subdivisão do traçado da rota. Ela determina o caminho a ser percorrido pelo leiturista, desde o seu ponto inicial até o final. Cada face recebe um número que mostra a seqüência do percurso do leiturista. Assim, na rota mencionada, o leiturista seguindo a seqüência dos números, iniciará seu percurso na face 01 e terminará na face 44.

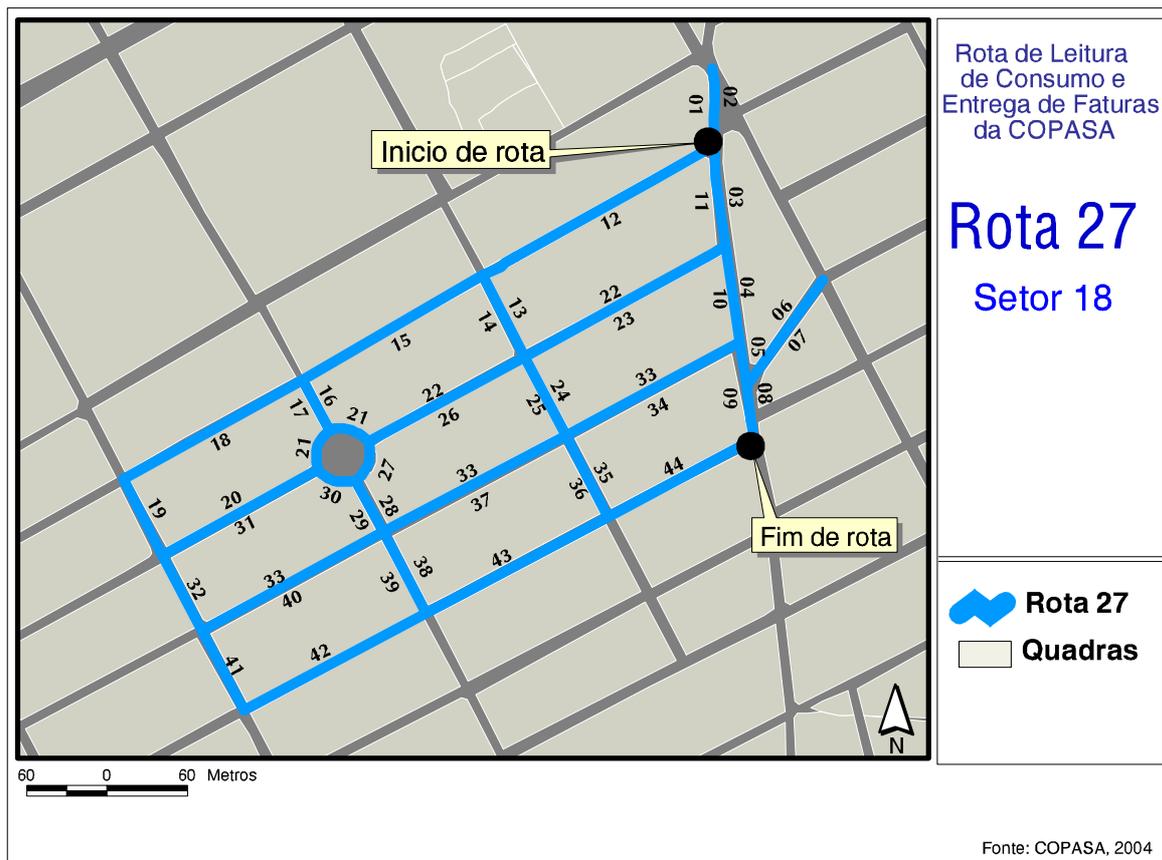


Figura 3.1.3 – Exemplo de uma rota de leitura de consumo e entrega de faturas da COPASA em Belo Horizonte.

Prosseguindo a definição da localização do imóvel, tomemos a face 08 como exemplo. Assim, teremos o número: 310620058-1827-08. Neste momento, já podemos identificar a UF, a localidade, o setor (parte da localidade), a rota (parte do setor) e a face que nos informa no mapa o logradouro, no qual se localizam os imóveis usuários. Agora, só falta localizarmos o imóvel cadastrado.

A figura 3.1.4 mostra um mapa com o detalhe da face 37 da rota 27 e do setor 18. Assim, tomemos como exemplo o imóvel de número 770, ele é identificado pela empresa através do número 060. Este número não foi escolhido por acaso. Percorrendo a face no sentido indicado (seqüência dos números das faces) os números de cadastramento dos imóveis (clientes) serão de acordo com a localização dos mesmos na seqüência (sentido) da face. O número oficial de escritura do imóvel está associado com um número seqüencial com o qual a empresa o identifica. Repare que o número de cadastramento cresce no sentido da face, de tal forma que o imóvel 060 vem depois do 040 e antes do 080. Se juntarmos este número aos anteriores, obteremos o *localizador* completo que identifica este imóvel. Assim, estará

identificada a Unidade Federativa 31 da localidade de Belo Horizonte 0620058, o setor 18, a rota 27, a face 37 e o imóvel 060. Então, o número 770 da Rua Dona Giusepella em Belo Horizonte, recebe o número *localizador* 310620058-1827-37060. Este número é único, isto é, nenhum outro imóvel usuário cadastrado pela empresa recebe número igual.

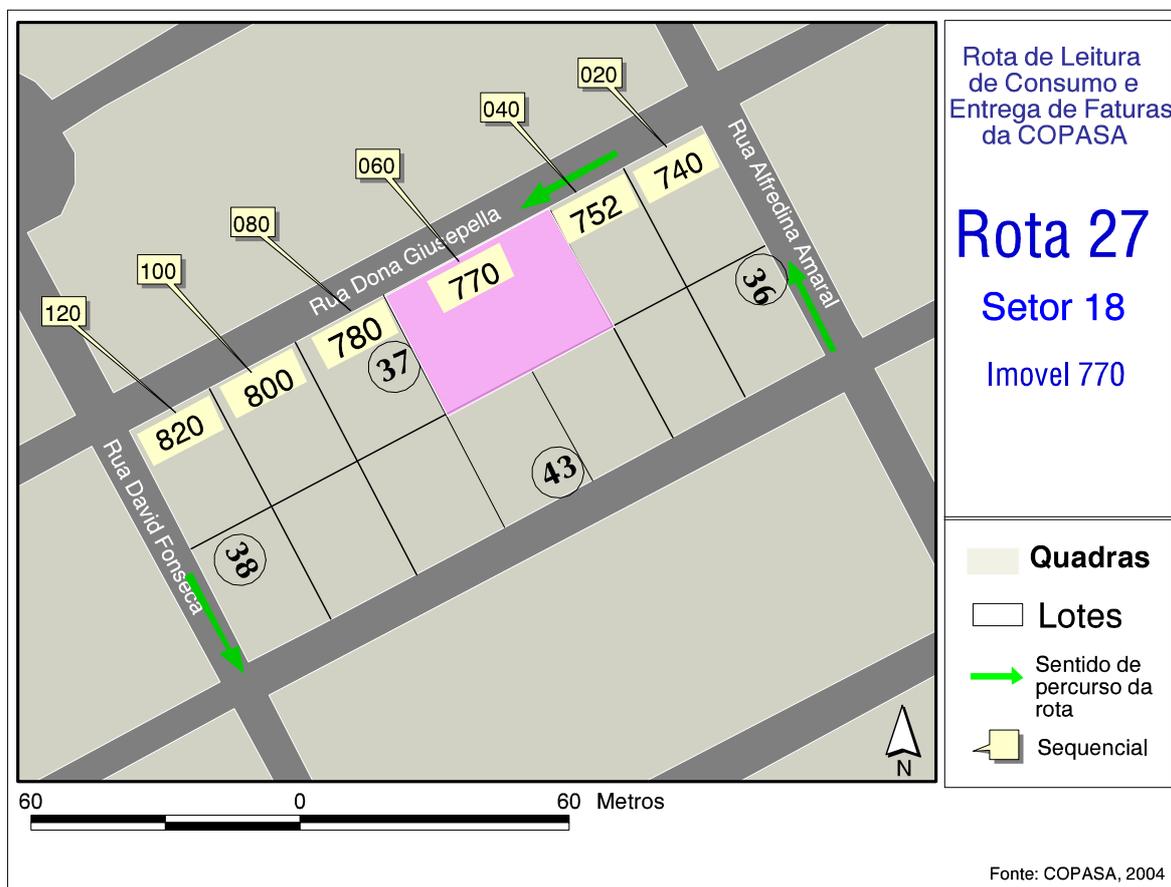


Figura 3.1.4 – Exemplo de localização de um imóvel pela metodologia de leitura de consumo e entrega de faturas da COPASA.

Entretanto, este não é o único número de identificação que aparece na fatura. Logo após o número do *localizador*, aparece um número que poderá ser composto por até 11 algarismos, que é chamado de matrícula do imóvel. Como o nome indica, este número permite a identificação do cliente no cadastro da empresa. O cliente, neste caso, é o imóvel e não o seu proprietário ou morador. O número de matrícula é extremamente útil para controle das informações comerciais e é determinado exclusivamente pelo sistema de processamento de dados. Estas informações, *localizador* e matrícula, permitem identificar facilmente o imóvel usuário e pesquisar no computador os registros cadastrais.

3.2 – Procedimentos Executados

A execução do trabalho seguiu a metodologia estabelecida inicialmente ⁵⁹, tendo sido realizada da seguinte forma:

3.2.1 – Fase de Especificação

Na primeira etapa desta fase ocorreu a obtenção do conhecimento através do levantamento de campo. Foram empregadas as técnicas de entrevistas e observação. Também foi observada em campo a metodologia de leitura de consumo e entrega de faturas da COPASA. Assim, a realidade foi estudada e o problema descrito detalhadamente. O resultado deste conhecimento está sintetizado neste capítulo, no item 3.1 - Introdução.

Também na especificação foi realizada uma revisão bibliográfica. Foram estudadas, em paralelo, obras referentes ao ambiente *GIS*, à técnica de *geocoding*, métodos e técnicas de desenvolvimento de sistemas de informação e banco de dados.

- **Modelagem do problema**

A etapa mais crítica da fase de especificação foi a modelagem do problema. Consistiu em representar a realidade através dos conceitos e recursos do ambiente *GIS*, aplicando a técnica de *geocoding* do *GIS* à metodologia de leitura de consumo e entrega de faturas da COPASA.

Partindo-se da premissa na qual as taxas de correspondência do processo de geocodificação se situam entre 25% a 75% para grandes bases de dados, procurou-se obter um método que viesse a produzir melhores resultados, bem como garantir que a atualização na base de clientes da COPASA (mantida pelo sistema comercial da empresa) se refletisse na nova base de dados georreferenciada de clientes a ser produzida e na atualização da base de dados de referência. Paralelamente, buscou-se adotar parâmetros para se atingir a melhor precisão possível de localização do imóvel.

⁵⁹ Ver item 2.4: “Metodologia e Técnicas Utilizadas”.

Considerando que os principais motivos para não se atingir uma taxa de correspondência de 100% estão relacionados às divergências entre a base de dados de referência e a base de dados de endereços alvo, a solução encontrada para este problema foi utilizar parte do localizador como atributo de relacionamento entre ambas as bases, isto é, todo o número do localizador, à exceção do número seqüencial que caracteriza individualmente o imóvel na face. A este código damos o nome de CÓDIGO DA FACE, pois identifica univocamente uma face no sistema comercial (SICOM).

Assim, a modelagem consistiu em criar para os trechos de logradouros atributos que identifiquem o CÓDIGO DA FACE. Como cada trecho de logradouro pode possuir duas faces (direita e esquerda), foram criados dois atributos: um atributo que contém o código do setor, da rota e da face para o seu lado direito (cd-srface-dir), e outro atributo que contém o código do setor, da rota e da face para o seu lado esquerdo (cd-srface-esq). Os atributos do localizador *unidade federativa, município, identificador de emancipação do município e dígito verificador* foram assinalados automaticamente para todos os trechos de um município em um único atributo (cd-município).

A figura 3.2.1 mostra os trechos A, B e C de um logradouro que são parte de um percurso da rota 01 do setor 15. Neste exemplo, e à esquerda destes trechos (de acordo com o sentido do trecho) encontramos as faces 01, 04 e 07. À direita, as faces 11 e 14. Desta forma, para o trecho A é assinalado para o atributo cd-srface-dir o valor 150114 e para o atributo cd-srface-esq o valor 150101. No trecho B, associamos a cd-srface-dir o valor 150111 e a cd-srface-esq o valor 150104. No trecho C, associamos a cd-srface-dir o valor 150111 e a cd-srface-esq o valor 150107, e assim sucessivamente, até completarmos toda a rota.

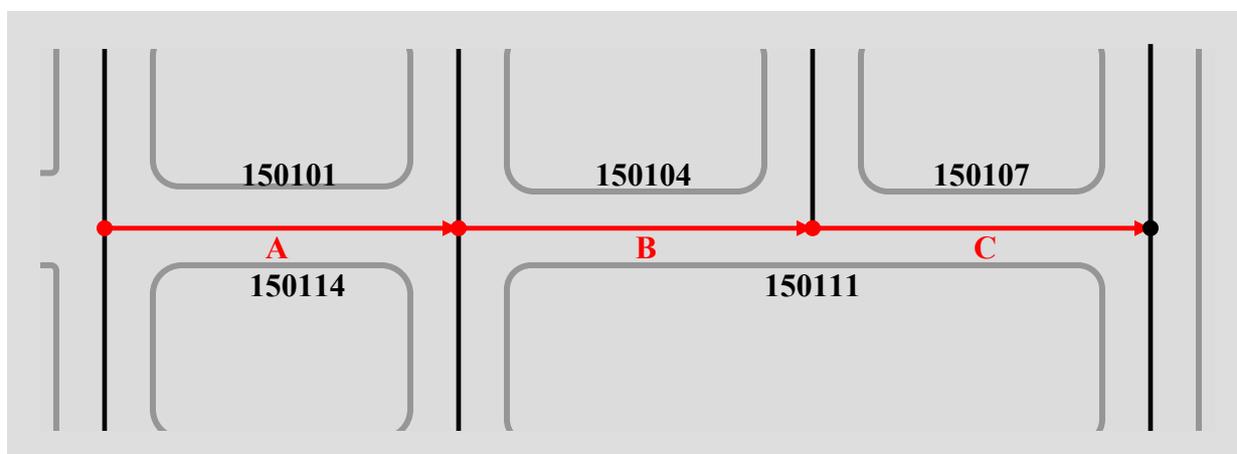


Figura 3.2.1 - Modelagem do problema na base de dados de referência.

Dessa forma, a solução utilizou a metodologia de leitura de consumo e entrega de faturas da COPASA e os recursos existentes no ambiente GIS. O trabalho foi dividido em três grandes etapas:

- Preparação da base de dados de referência
- Obtenção da base de dados de endereços alvo
- Geocodificação dos clientes COPASA

- **Preparação da base de dados de referência**

Esta etapa consistiu basicamente em utilizar os recursos do GIS para a atualização da base de dados de referência, ou seja, do *layer* trechos. Buscou-se, em primeiro lugar, utilizar ferramentas que possibilitassem a localização geográfica da região a ser trabalhada. Foram especificadas funções de localização baseadas no recurso do GIS para trabalhar com dados alfanuméricos associados às feições geográficas (linha, ponto e polígono). Foi especificado, também, a localização por endereço através do recurso de *geocoding* do Arc/Info (*Address Matching*), baseado nos atributos já existentes no *layer* de trechos⁶⁰ obtido junto à prefeitura de Belo Horizonte.

Para a preparação da base de dados de referência é necessário identificar os trechos de logradouros pertencentes a uma determinada rota de leitura de consumo utilizando qualquer uma das funções de localização especificadas. Para tal, o operador deverá possuir em mãos o croqui da rota de leitura a ser cadastrada (figura 3.2.2) e localizá-la no mapa em tela produzido pelo SIURB. O operador seleciona também a escala de trabalho que desejar utilizando as ferramentas de *zoom* do GIS.

⁶⁰ Ver tabela 3.1.1: “Atributos básicos dos trechos de logradouros”.

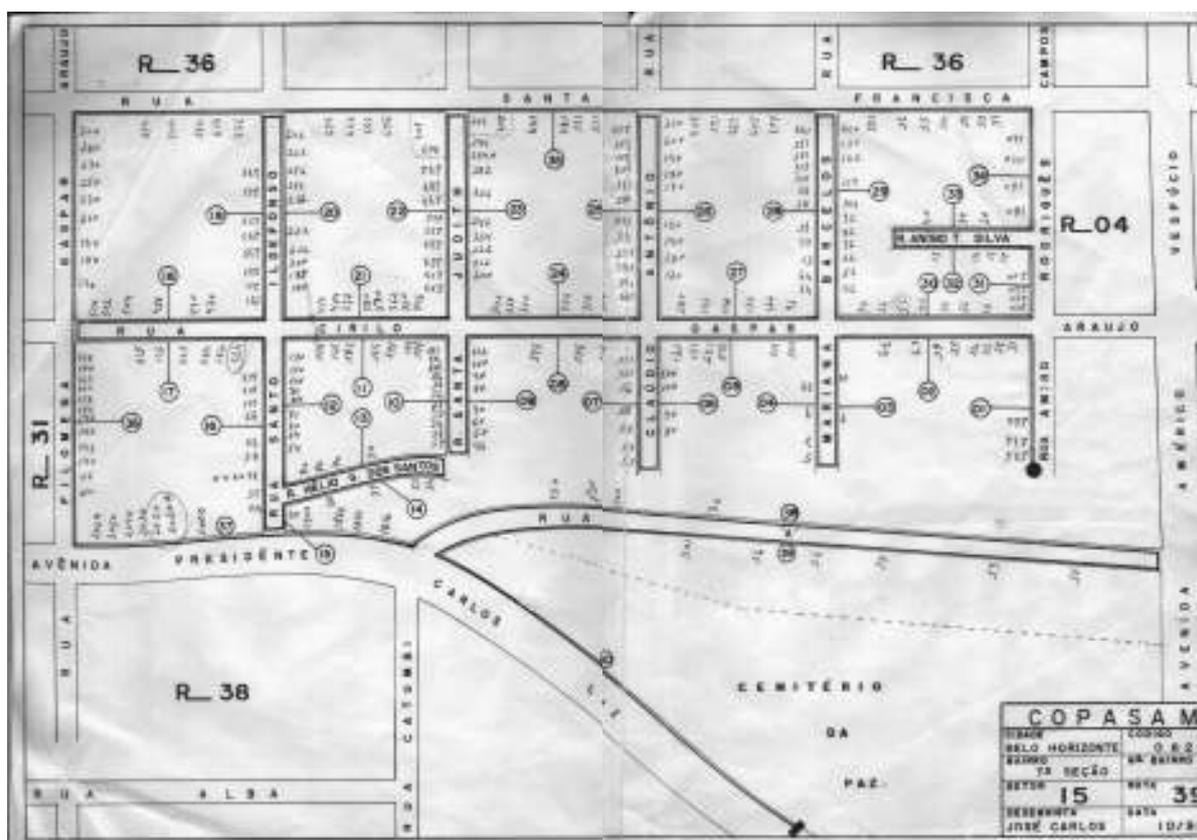


Figura 3.2.2 – Croqui da rota de leitura 39 do setor 15.⁶¹

Uma vez identificados os trechos, o operador percorre a rota cadastrando o setor, rota e face à direita e à esquerda para cada trecho de logradouro de acordo com o croqui existente (figura 3.2.3). Caso haja uma divergência entre o croqui e a base de trechos de logadouros, é necessária uma conferência em campo.

⁶¹ Este croqui apresenta a anotação da numeração dos imóveis individualmente. Entretanto, essa numeração não é atualizada e não está presente em todos os croquis de rota existentes.

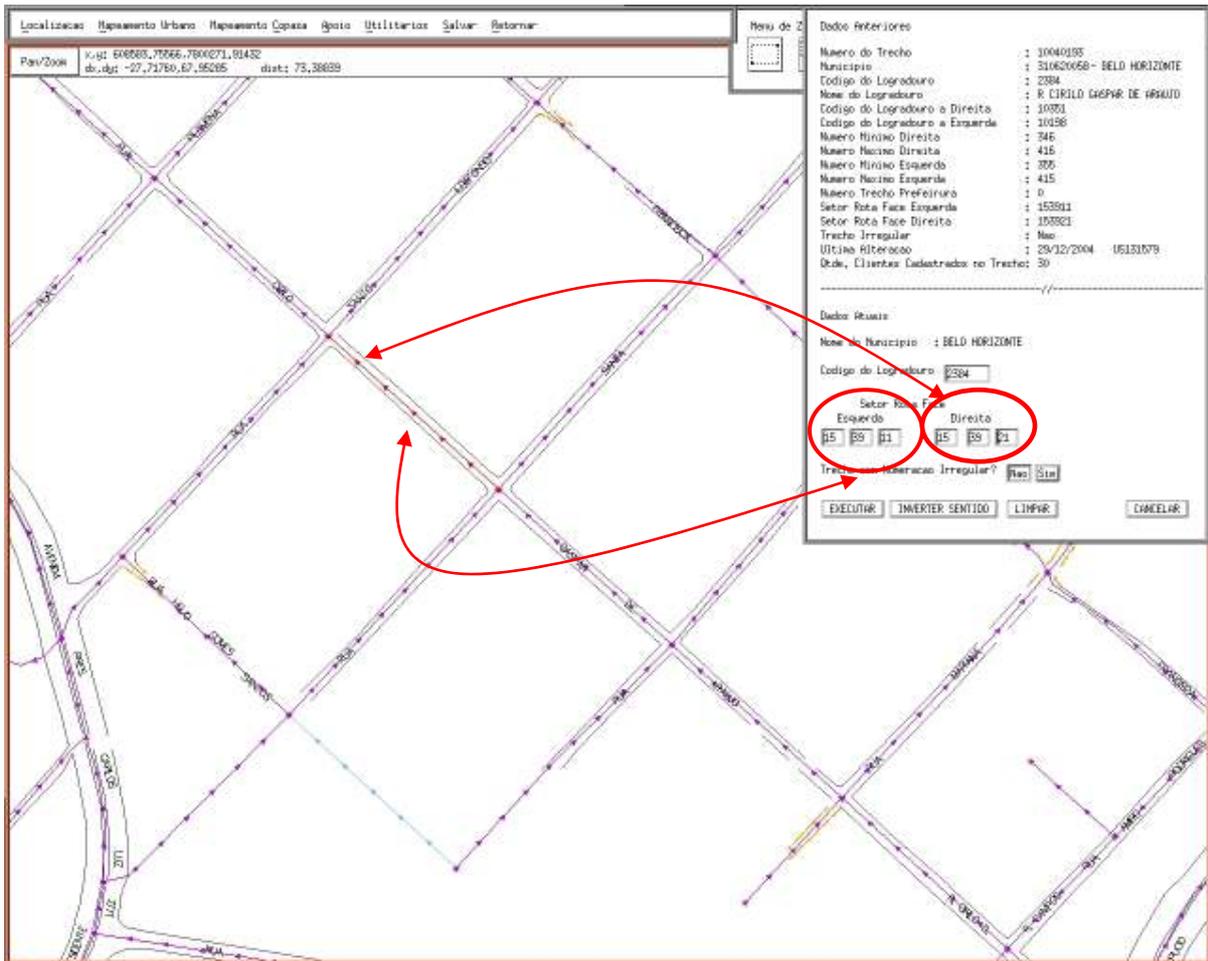


Figura 3.2.3 – Exemplo do processo de cadastramento do setor-rota-face nos trechos de logradouros (setor 15 rota 39).⁶²

⁶² Fonte: Imagem capturada da tela do sistema SIURB.

Cabe ressaltar que é necessário haver uma consistência do sentido do trecho de logradouro. Caso o sentido do trecho esteja incorreto, haveria um posicionamento incorreto do imóvel no lado errado do logradouro. Ao cadastrar a rota de leitura, o operador também verifica o sentido de crescimento da numeração ao longo do logradouro para corrigir, se necessário, o sentido dos trechos. Dessa forma, o sentido dos trechos de logradouro deve acompanhar o sentido de crescimento da numeração dos imóveis. Caso haja algum engano ou esquecimento por parte do operador, o SIURB deverá indicar os trechos com sentido incorreto. Para tal, são utilizadas as informações topológicas atribuídas pelo *software* de GIS no modelo arco-nó, ou seja, não poderá haver um nó que seja um nó de início ou um nó de fim de dois arcos distintos e que pertençam ao mesmo logradouro (figura 3.2.4).

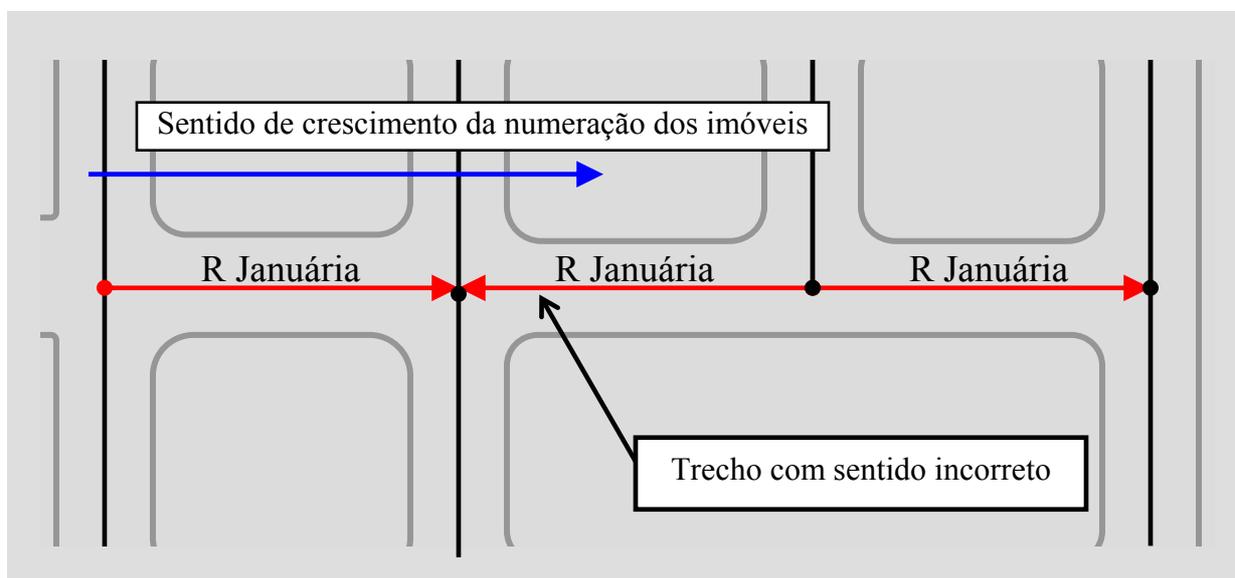


Figura 3.2.4 – Exemplo de trecho de logradouro com sentido incorreto.

• Obtenção da base de endereços alvo

Uma vez realizado o cadastramento das rotas conforme indicado no item anterior, é necessário obter a base de endereços alvo, que neste caso, é a base de clientes mantida pelo sistema SICOM. Como os sistemas SIURB e SICOM possuem ambientes de computadores distintos e estão interligados em rede de computadores, foi utilizado o serviço de FTP ⁶³ da rede TCP-IP ⁶⁴ da COPASA para a geração da base de endereços do SIURB.

⁶³ File Transfer Protocol (Protocolo para a transferência de arquivos entre computadores em rede).

⁶⁴ Transport Control Protocol – Internet Protocol (Protocolo para a comunicação entre computadores em rede).

Este processo é diário e realizado automaticamente sem a interferência do operador. Compõe-se das seguintes etapas:

- Identificação das faces já cadastradas nos trechos de logradouros e ainda não processadas para a seleção dos clientes a serem geocodificados. Estas rotas são obtidas após o processo de cadastramento descrito no item anterior. Um arquivo é gerado contendo somente os trechos que foram atualizados, sendo posteriormente enviado para o ambiente mainframe onde deverá ser processado.
- Seleção dos clientes a serem geocodificados. A base de dados de clientes do SICOM é acessada e são selecionados todos os clientes que pertencem às faces identificadas e contidas no arquivo enviado. Um arquivo com os clientes selecionados é gerado e enviado para o ambiente GIS.

Também nesta etapa é necessário verificar a existência de algum erro de cadastramento do operador. Caso o operador inclua alguma face (setor, rota e face) para a qual não existam clientes cadastrados no SICOM, o sistema deverá gerar um relatório de consistência para que esta ocorrência possa ser verificada pelo operador. Na verdade, esta situação não é necessariamente um erro, uma vez que podem ocorrer faces sem que existam clientes associados.

• **Geocodificação dos clientes COPASA**

Uma vez enviados os clientes obtidos junto ao sistema comercial para o ambiente GIS, pode ser dado início à geocodificação destes endereços para a produção do *layer* clientes (arquivo geográfico com os endereços geocodificados). Para tal, foi utilizada a funcionalidade de *geocoding* do *software* de GIS Arc/Info versão 8.2. Esta funcionalidade, de acordo com o processo exposto no item 2.2.3, normalmente é executada uma única vez para o conjunto de todos os endereços alvo contra toda a base de dados de referência, gerando um arquivo de pontos geocodificados e um arquivo com os endereços rejeitados. Entretanto, na metodologia proposta, este processo é repetido tantas vezes quantas forem as faces a serem processadas.

Utilizando as técnicas de programação de computadores e o recurso do GIS em trabalhar com atributos alfanuméricos, o processo de geocodificação é executado para um subconjunto do arquivo de endereços alvo contra um subconjunto do arquivo de referência. O atributo de seleção destes subconjuntos é o código da face. Assim, os comandos de *geocoding* do

Arc/Info são executados para todos os endereços de uma única face contra uma base de referência reduzida a um único trecho de logradouro que representa esta face (figura 3.2.5).

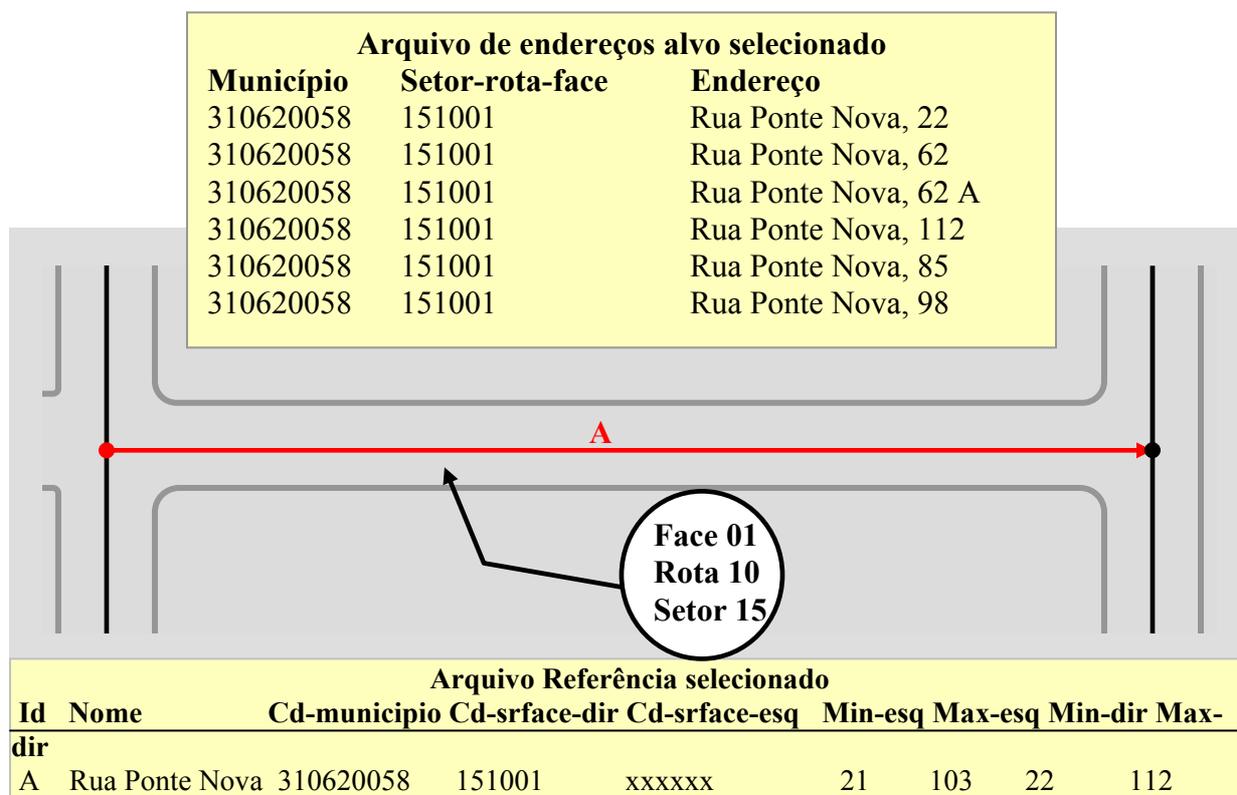


Figura 3.2.5 – Ilustração do modelo de *geocoding* adotado que trabalha com subconjuntos dos arquivos de endereços alvo e de referência.

Durante este processo, entretanto, alguns artifícios são utilizados para solucionar o problema da numeração irregular e da rejeição dos endereços pela desatualização das faixas de numeração da base de referência, a saber: a **criação de uma numeração fictícia**, a **criação de um nome de logradouro fictício** e o **ajuste dos fatores de afastamento e encurtamento**.

Criação de uma numeração fictícia.

Para cada endereço da face processada é criado um número de imóvel fictício baseado no número seqüencial do localizador. Obedecida a seqüência indicada por este número e considerando a quantidade total de endereços da face, é gerado automaticamente um número fictício.

Para faces localizadas à direita do trecho de logradouro, inicia-se a numeração pelo número 02 (dois) ⁶⁵. Os próximos endereços da face recebem um número fictício do anterior acrescido de 50. Exemplo: 02, 52, 102, 152, etc.

Para faces localizadas à esquerda do trecho de logradouro, inicia-se a numeração pelo número 01 (um). Os próximos endereços da face recebem um número fictício do anterior acrescido de 50. Exemplo: 01, 51, 101, 151, etc.

Este artifício visa solucionar dois problemas: a numeração irregular do trecho de logradouro e a geocodificação de endereços de mesmo número do imóvel (várias edificações no mesmo lote). Caso o trecho venha a possuir uma numeração irregular, não seria possível utilizar o número do imóvel, uma vez que este caso não é modelado pelo padrão de geocodificação utilizado. Entretanto, a numeração do imóvel não é perdida, já que ela é um atributo que é incorporado à feição de ponto criada para representar o endereço. Esta numeração fictícia só é utilizada para o processo de geocodificação e depois descartada.

A numeração fictícia também é utilizada para criar um modelo de representação dos imóveis de mesmo endereço, mas com complementos diferentes. Como o modelo de geocodificação não trata esta situação (os imóveis de mesmo endereço teriam obrigatoriamente a mesma coordenada), foi utilizada a numeração fictícia para construir um modelo que gere coordenadas próximas para endereços iguais. Neste caso, quando se identifica dois ou mais imóveis de mesmo endereço, o número fictício dos demais imóveis de mesmo endereço recebem o número fictício do anterior acrescido de 02. Exemplo: 01, 51, 53, 55, 101, 151, onde os números fictícios 51, 53 e 55 correspondem a três imóveis de mesmo endereço. Este artifício faz com que no processo de interpolação estes pontos tenham coordenadas próximas uns dos outros ao longo do trecho.

Uma vez estabelecidos os números fictícios para todos os endereços da face, o trecho de logradouro é atualizado com a faixa de numeração representada pelo número fictício inicial e final existente no arquivo de endereços da face.

⁶⁵ Por convenção do processo de geocodificação, os imóveis com numeração par são posicionados à direita dos trechos de logradouro.

Este artifício, todavia, irá provocar uma distribuição uniforme dos endereços ao longo do trecho de logradouro, o que pode não representar a realidade em campo. Entretanto, a eliminação do problema da numeração irregular aumenta a taxa de correspondência e a precisão posicional atingida atende as necessidades de informação da COPASA.

Criação de um nome de logradouro fictício.

Se o processo de geocodificação tivesse início neste ponto, poderíamos enfrentar situações de rejeição de endereços por incompatibilidade de endereços entre o arquivo de endereços alvo e o arquivo de referência, devida a alguma diferença no nome do logradouro. Neste ponto, outras causas de rejeição já estão descartadas, uma vez que a numeração do arquivo de referência foi gerada durante o processo e não há como haver múltiplos trechos de logradouro para o mesmo endereço (somente uma única face está sendo processada por vez). Assim, já que a correspondência entre os endereços alvo e o arquivo de referência está garantida pela etapa anterior de preparação da base de dados de referência, nada impede que se utilize um nome de logradouro fictício em ambos os arquivos para que não haja nenhuma possibilidade de rejeição de endereços. Desta forma, a todos os endereços do arquivo de endereços alvo e ao trecho de logradouro do arquivo de referência é atribuído momentaneamente um nome de logradouro fictício (neste caso, foi simplesmente atribuído o caractere “A”).

Ajuste dos fatores de afastamento e encurtamento.

A maioria dos *softwares* de GIS possui parâmetros de ajuste da geocodificação para melhor posicionamento das coordenadas calculadas.⁶⁶ No caso do *software* Arc/Info, o fator de afastamento é chamado de *offset distance* e o fator de encurtamento, *squeeze factor*. Neste trabalho, foram adotados três valores de afastamento de acordo com o tipo de logradouro de cada trecho (tabela 3.2.1).

Para o fator de encurtamento, como a metodologia de *geocoding* adotada processa cada face por vez, o valor utilizado foi calculado pela avaliação do resultado da razão entre o valor 2400 pelo comprimento do trecho (tabela 3.2.2). Estes valores foram assinalados empiricamente pela observação dos melhores resultados obtidos.

⁶⁶ Ver item 2.2.3: “O processo de geocodificação de endereços” para uma descrição destes parâmetros.

	Tipos de logradouros		
	Beco	Avenida	Demais tipos
<i>offset distance</i>	5	14	10

Tabela 3.2.1 – Fatores de afastamento utilizados.

	$X = 2400 / \text{comprimento do trecho}$		
	$40 < X \leq 55$	$55 < X \leq 70$	$40 \geq X > 70$
<i>squeeze factor</i>	40	30	20

Tabela 3.2.2 – Fatores de encurtamento utilizados.

Agora, então, o processo de geocodificação pode ser realizado, tendo como entrada os dois arquivos devidamente preparados. Assim, é esperado que o processo ocorra resultando em uma correspondência de 100% para os endereços alvo, uma vez que todas as possibilidades de rejeição foram contornadas no processo de preparação da base de referência.

- **Elaboração do Modelo Ambiental**

Nesta etapa foi elaborada a matriz de eventos, o diagrama de contexto e a declaração dos objetivos do sistema que podem ser apreciados no anexo 1.

- **Elaboração do Modelo Comportamental**

Nesta etapa foi elaborado o diagrama de fluxo de dados preliminar e o diagrama de entidades-relacionamentos que podem ser apreciados no anexo 2.

O aplicativo possui as seguintes funcionalidades básicas:

- *Localização da área a ser trabalhada:*

Refere-se aos processos de identificação do local a ser trabalhado para a atualização dos trechos de logradouros e visualização dos clientes georreferenciados. Consta de:

- Localização por endereço;
- Localização fonética de endereço;
- Localização por código de localizador (setor-rota-face);
- Localização por logradouro.

- *Geocodificação dos clientes da COPASA:*

Refere-se aos processos de geocodificação dos clientes da COPASA. Consta de:

- Atualização dos trechos de logradouros com o código da face;
- Obtenção dos clientes de todas as faces cadastradas;
- Geocodificação dos clientes;
- Atualização dos clientes geocodificados quando do remanejamento de localizador, inclusão e exclusão de clientes no sistema comercial.

- *Controle de Zoom:*

Refere-se às facilidades de operações de *Zoom* (*Zoom in, Zoom out, Zoom by box, etc.*).

- *Controle de visualização:*

Permite a escolha dinâmica dos níveis da base de dados a serem visualizados, incluindo os textos gráficos.

3.2.2 – Fase de Projeto Físico

- **Projeto da Arquitetura Física de Implementação**

Nesta etapa foi elaborado o projeto da arquitetura de implementação do sistema, ou seja, o projeto do *hardware* e *software* necessário ao funcionamento do aplicativo. O aplicativo foi implementado em rede na seguinte plataforma:

Hardware:

1 máquina RISC - IBM RS/6000 com o sistema operacional AIX.

Softwares:

ARC/INFO, NETWORK e ARCSTORM;

Linguagem AML.

- **Projeto Físico das Bases de Dados**

O ambiente Arc/Info na plataforma implementada utiliza o formato chamado *coverage* para armazenar os níveis de informações geográficas. Como o Sistema SIURB foi desenvolvido no ambiente Arc/Info, nesta etapa foi elaborado o projeto das *coverages*, dos arquivos INFO e do banco de dados ARCSTORM do ambiente. Este projeto é mostrado no anexo 3.

- **Estruturação Física do Sistema**

Nesta etapa foi gerado o Diagrama Físico do Sistema (anexo 4) que retrata a seqüência de execução dos programas do sistema.

- **Especificação das Entradas e Saídas**

Nesta etapa foram elaborados os *lay-outs* de telas, relatórios e mapas do sistema. O item 3.3 - Produtos Finais apresenta alguns destes produtos.

- **Definição de Programas (algoritmos)**

Esta etapa consistiu da definição e elaboração dos algoritmos utilizados no sistema.

Considerando que a metodologia de geocodificação proposta executa os comandos de *geocoding* do Arc/Info para cada face das rotas de leitura de consumo e entrega de faturas da COPASA, este programa foi desenvolvido para ser executado no horário noturno para não comprometer a desempenho geral do sistema. Assim, durante a etapa de cadastramento das rotas (preparação da base de dados de referência) são identificados quais os trechos de logradouros que receberam o código da face para serem processados pelo programa de geocodificação noturno.

- **Elaboração e Teste de Programas**

Nesta etapa os algoritmos definidos foram codificados na linguagem do ambiente ARC/INFO, ou seja a AML - ARC MACRO LANGUAGE. O anexo 5 mostra o código do programa que executa a geocodificação dos clientes.

- **Teste Integrado do Sistema**

Nesta etapa foi escolhido todo um setor de leitura de consumo e entrega de faturas da COPASA para a realização do teste integrado. O setor escolhido foi o setor 15 por contemplar todas as restrições do problema.

Durante os testes foram feitos alguns ajustes necessários, inclusive da determinação dos melhores parâmetros de posicionamento (*offset distance, squeeze factor*). O resultado do teste encontra-se relacionado no item 3.4 – Análise e validação dos resultados.

3.3 – Produtos Finais

O produto final gerado por este trabalho corresponde ao aplicativo computacional SIURB, operando conforme as regras e restrições impostas pelo problema. Foram gerados 54 programas, que implementam os processos necessários à geocodificação dos clientes da COPASA. Assim, o próprio *software* em condições adequadas de operação e funcionamento, corresponde ao resultado final do trabalho.

Paralelamente à produção do *layer* clientes, ou seja, à geocodificação dos clientes da COPASA, o SIURB proporciona, também, uma atualização dos atributos de faixa de numeração do arquivo de referência (*layer* trechos), baseando-se na numeração dos imóveis do arquivo de clientes enviado pelo sistema comercial para o processo de geocodificação. A atualização destes atributos é, portanto, um resultado adicional que se obtém à medida que se constrói a base de clientes georreferenciados.

Alguns produtos podem ser visualizados através dos relatórios de consistência impressos e mapas gerados na modalidade *on-line*.

3.3.1 - Mapas de rota

Apresenta os trechos que correspondem a uma determinada rota com os seus respectivos códigos de face à direita e à esquerda (figura 3.3.1).



Figura 3.3.1 – Mapa da rota 54 setor 15.⁶⁷

⁶⁷ Fonte: Imagem capturada da tela do Sistema SIURB.

3.3.2 - Mapas dos clientes georreferenciados

Após a geocodificação, os endereços são armazenados em um banco de dados de clientes georreferenciados no SIURB e disponíveis também aos demais sistemas de informações geográficas da COPASA. A figura 3.3.2 ilustra um mapa produzido pelo SIURB que apresenta toda uma rota de leitura com os respectivos clientes cadastrados.

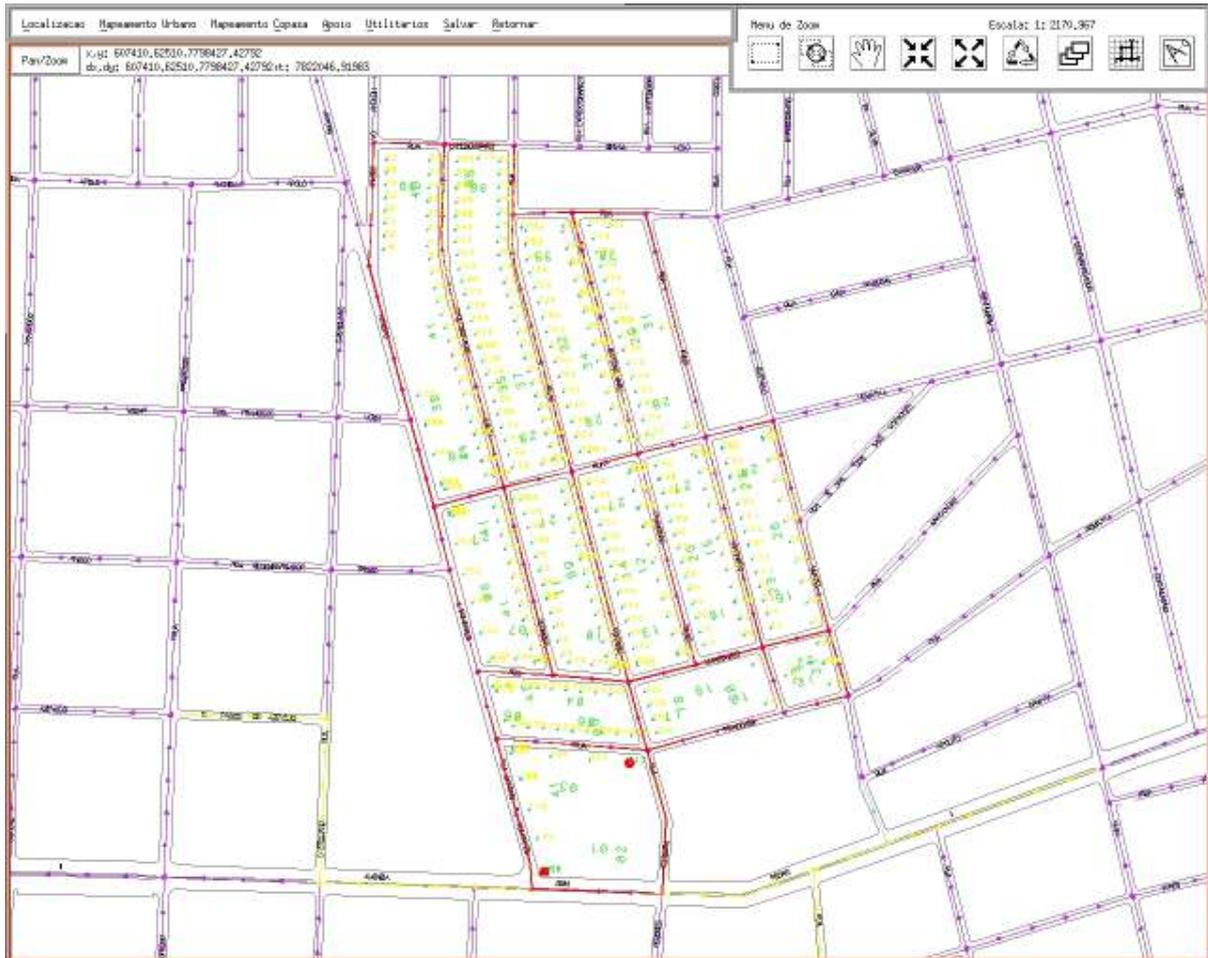


Figura 3.3.2 – Mapa dos clientes geocodificados da rota 01 setor 15.⁶⁸

⁶⁸ Fonte: Imagem capturada da tela do Sistema SIURB.

3.3.3 - Relatórios de consistência

Os relatórios de consistência produzidos pelo SIURB são utilizados para a verificação dos dados de entrada do sistema, compreendendo os relatórios:

- “Relação de trechos com setor-rota-face em duplicidade” (figura 3.3.3), utilizado para localizar os trechos de logradouro que possuem códigos de face em duplicidade e fora do previsto;
- “Relação de logradouros com sentidos invertidos” (figura 3.3.4), utilizado para verificar os logradouros cujos sentidos dos trechos contém alguma irregularidade;
- “Setor-rota-face não encontrado no Sistema Comercial”, utilizado para indicar os códigos de faces cadastrados no SIURB, mas que não estão cadastrados no SICOM;
- “Logradouros com setor-rota-face incompatíveis”, utilizado para indicar uma discordância do logradouro cadastrado no SIURB e no SICOM para o código da face cadastrado.

RELACAO DE TRECHOS COM SETOR-ROTA-FACE EM DUPLICIDADE										
Distrito: SPIN Data/Hora: 23/05/2005 - 21.00.48										
Pag.: 1										
MUNICIPIO	SRFACE	TRECHO1	TRECHO2	ERROS	LOGR1	LOGR2				
310620058	151004	10040652	10040649	4	2514	99999	R	CONDE DE VALADARES		
310620058	151610	10039425	10039703	3	9112	9165	R	PIGMATITA / R PIRACICABA		
310620058	156618	10055558	10055559	2	8772	9999		AL PALMAS / R ROSALINA BANDEIRA		
ERRO1: Faces Direita/Esquerda em duplicidade para o mesmo logradouro										
ERRO2: Face Direita em duplicidade para logradouros diferentes										
ERRO3: Face Esquerda em duplicidade para logradouros diferentes										
ERRO4: Faces Direita/Esquerda em duplicidade para logradouros diferentes										
ERRO5: Faces duplicadas não subseqüentes										

Figura 3.3.3 – Relatório de consistência do SIURB contendo as faces cadastradas em duplicidade.

RELACAO DE LOGRADOUROS COM SENTIDOS INVERTIDOS									
Distrito: spin Data/Hora: 23/05/2005									
MUNICIPIO	PREFEITURA	COPASA	NU-TRECHO	SRFACDIR	SRFACESQ	TIPO	NOME	LOGRADOURO	
310620058	0041899	007215	10039764	091310	151426	RUA	LUNARDI		
310620058	0067480	011283	10039854	151841	000000	RUA	TEOTONIO MACIEL		
310620058	0301017	001093	10040610	152438	152439	BEC	ASSOCIACAO		
310620058	0300692	000120	10040515	156625	156621	ALA	ACUCENA		

Figura 3.3.4 – Relatório de consistência do SIURB contendo os trechos de logradouro com sentido invertido.

3.4 – Análise e Validação dos Resultados

A análise dos resultados e sua validação basearam-se, principalmente, na análise comparativa dos resultados observados da geocodificação pela metodologia habitual (diretamente pelo endereço dos imóveis cadastrados pela COPASA), com os resultados observados da geocodificação pela abordagem proposta. Para efeito desta análise, as seguintes observações foram realizadas:

1. A taxa de geocodificação (*address-matching rate*), ou seja, o percentual dos endereços que foram geocodificados a partir do universo de endereços do contexto de validação (setor 15);
2. O percentual de endereços corretamente posicionados dentro dos limites da quadra. Para esta avaliação será utilizado o nível de limite de quadras fornecido pela Prefeitura de Belo Horizonte;
3. O tratamento dado por cada metodologia ao problema da numeração irregular, pela observação de um exemplo real conhecido onde se verifica esta situação.

O universo de endereços correspondentes ao setor 15 é de 23.871 endereços. Estes endereços foram obtidos junto ao sistema comercial da COPASA e correspondem ao arquivo de endereços alvo utilizado em ambas as metodologias. Para a base de referência foi utilizado o arquivo georreferenciado de trechos de logradouros obtido junto à prefeitura de Belo Horizonte, através de seu órgão de informática, a Prodabel.

3.4.1 – Observação da Taxa de Geocodificação

Metodologia habitual

A figura 3.4.1 abaixo retrata as estatísticas do processo de geocodificação utilizando a metodologia habitual. Foi utilizado o processo de geocoding do software ArcGis, versão 8.3. Nota-se a obtenção de uma taxa de geocodificação de 63%.

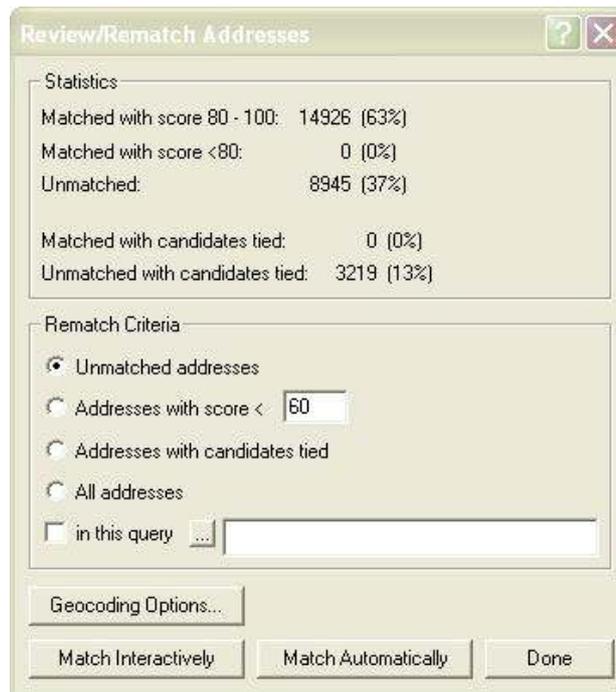


Figura 3.4.1 – Estatística do processo de geocoding do ArcGis 8.3 para o setor 15 dos clientes da COPASA.

A análise dos endereços não geocodificados aponta para as seguintes causas para a existência do percentual de 37% de endereços não correspondidos:

1. Ausência de faixas de numeração na base de referência. Alguns eixos de logradouros não possuem os valores das faixas de numeração, ou não apresentam uma faixa que inclua a numeração pretendida. Este problema é devido à desatualização da base de referência que, entre outras razões, não acompanha o crescimento urbano;
2. Sobreposição de faixas de numeração, ou seja, faixas de numeração máxima e mínima, à esquerda e à direita, que se sobrepõe, ocasionando múltiplas possibilidades de geocodificação. Neste caso, o software foi configurado para rejeitar o endereço, uma vez que não se pode precisar com exatidão a qual trecho de logradouro pertence o imóvel;
3. Incompatibilidade de nome de logradouros entre os dois bancos de dados, ou seja, entre a base de eixos de logradouros de referência e o arquivo de endereços. Isto é devido, principalmente, pela existência de duas fontes distintas dos dados: a prefeitura de Belo Horizonte, por um lado, e a COPASA, por outro;

4. Base de dados de eixos de logradouros incompleta. Alguns logradouros não estão representados na base de referência, principalmente nas regiões de favelas.

Metodologia proposta

Pela metodologia proposta, dos 23.871 endereços do setor 15, foram geocodificados 23.247 endereços, correspondendo a uma taxa real de geocodificação de 97,4%. Entretanto, a taxa de geocodificação efetiva foi de 100%, conforme esperado,⁶⁹ uma vez que o arquivo de endereços alvo enviado pelo sistema comercial continha 23.247 endereços. Os 624 endereços restantes não foram geocodificados por problemas na identificação das faces no layer de trechos de logradouros, o que levou o SIURB a não solicitar os endereços junto ao sistema comercial. A análise dos endereços não geocodificados aponta para os seguintes problemas:

1. Base de dados de eixos de logradouros incompleta. Alguns logradouros não estão representados na base de referência, principalmente nas regiões de favelas. Um exemplo deste caso foi identificado na rota 49, cujo croqui está diferente do arquivo de trechos de logradouros de referência (figura 3.4.2). Neste caso, faz-se necessário uma conferência em campo;
2. Erros no lançamento do código de setor-rota-face. Estes erros são devidos aos erros humanos na entrada de dados do SIURB, pela omissão de setor-rota-face ou erros de digitação.

⁶⁹ Ver item 3.2.1 – Fase de Especificação

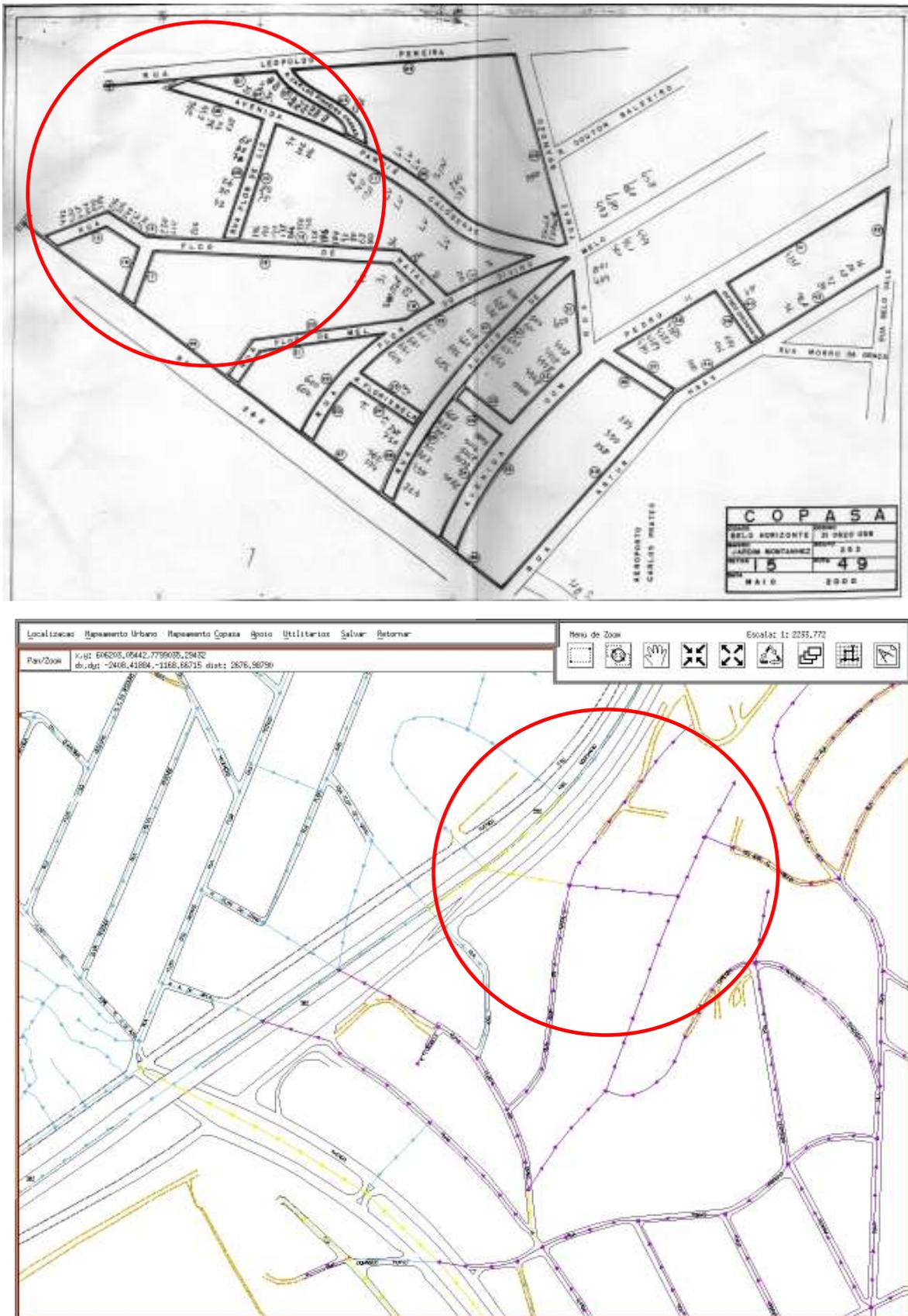


Figura 3.4.2 – Exemplo de uma região cujos trechos de referência não correspondem ao croqui da rota (rota 1549).⁷⁰

⁷⁰ O círculo vermelho representa a região em desacordo entre o croqui e os trechos de referência visualizados na tela do sistema SIURB.

3.4.2 – Observação do Posicionamento Dentro dos Limites de Quadras

Este item foi incluído para se analisar a distribuição espacial dos endereços geocodificados. Embora o ponto geográfico seja somente um modelo de representação do imóvel (uma melhor representação seria o polígono do lote ou da edificação), é desejável que este ponto esteja posicionado dentro dos limites da quadra. Para este levantamento, foi efetuada uma pesquisa espacial dos endereços contidos nos polígonos de quadras e utilizado o software ArcGis 8.3. O resultado mostrou que dos 23.247 endereços geocodificados, 21.988 endereços estavam contidos nos polígonos das quadras, correspondendo a aproximadamente 95% do total. Portanto, cerca de 5% dos endereços estão situados fora dos limites das quadras, evidenciando um posicionamento indesejado dos endereços. Uma avaliação dos mapas que mostram estes resultados indesejados indica três situações:

- Uma face pode ser formada por dois ou mais trechos de um mesmo logradouro. Se estes trechos estão separados por um outro logradouro, a metodologia adotada une estes trechos e realiza o processo de geocodificação como se não houvesse esta separação, podendo ocorrer que um endereço seja posicionado no meio da rua (figura 3.4.3). Esta situação, embora real, é devida a um erro na construção da rota, que deveria, pelo padrão adotado pela COPASA, mudar de número de face ao atravessar um logradouro;

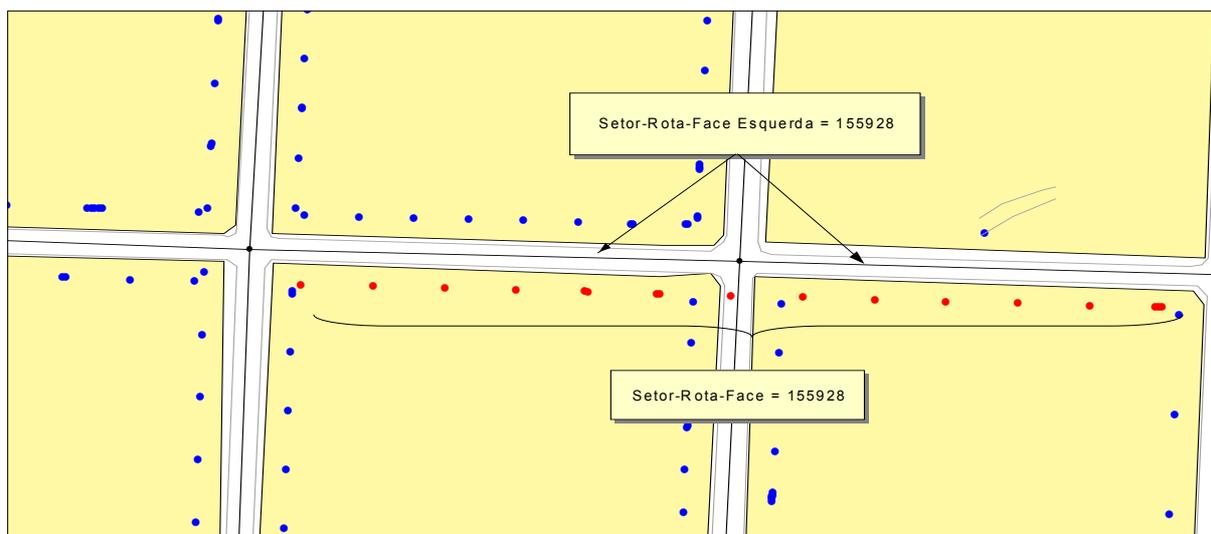


Figura 3.4.3 – Erro de posicionamento devido a um erro construtivo da rota. ⁷¹

⁷¹ Fonte: Imagem capturada da tela do mapa produzido pelo software ArcGis 8.3.

- Uma face pode corresponder a uma esquina que tenha um ângulo agudo suficientemente acentuado para que a quadra esteja situada a uma distância maior do que o fator de encurtamento adotado. A figura 3.4.4 mostra os números 242 (Av Ressaca) e 121 (Rua Dona Ana) posicionados fora das quadras. Um fator de encurtamento mais adequado corrigiria esta situação;

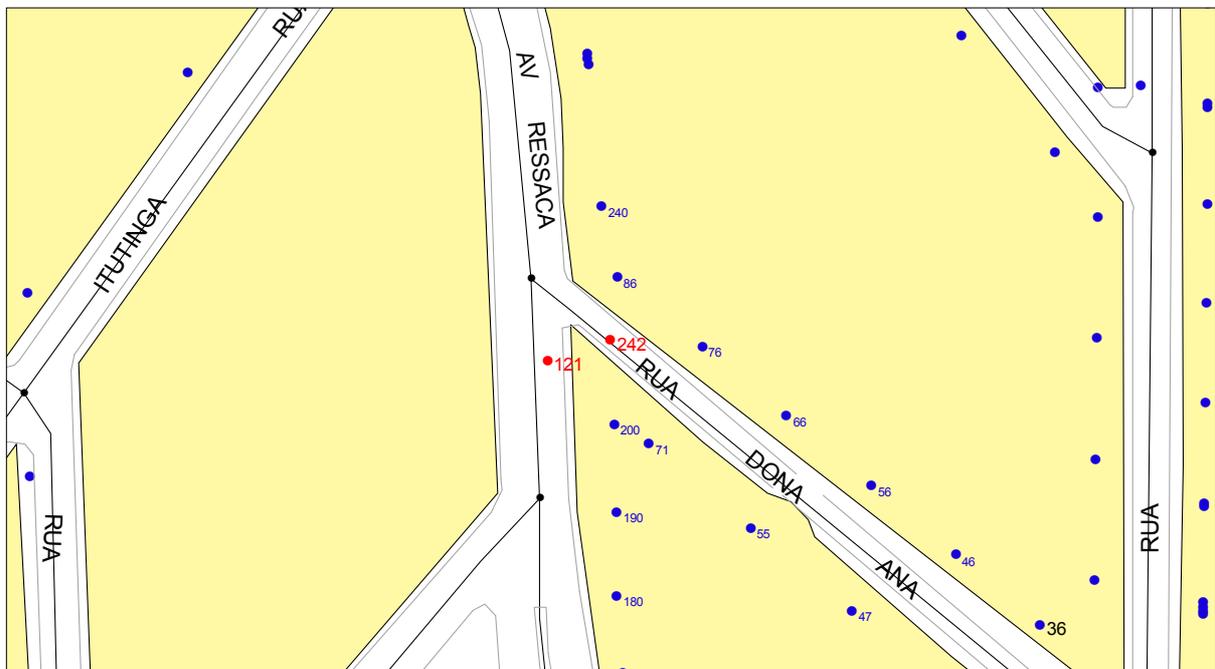


Figura 3.4.4 – Erro de posicionamento devido ao ângulo agudo acentuado da esquina.⁷²

- Os fatores de afastamento adotados não se aplicam a todos os casos. Quando se trata de uma avenida cuja largura ultrapassa o fator adotado de 14 metros (tabela 3.2.1) conforme mostra a figura 3.4.5, os pontos geocodificados em vermelho são posicionados à 14 metros do trecho de logradouro correspondente, mas fora dos limites da quadra.

⁷² Fonte: Imagem capturada da tela do mapa produzido pelo software ArcGis 8.3.

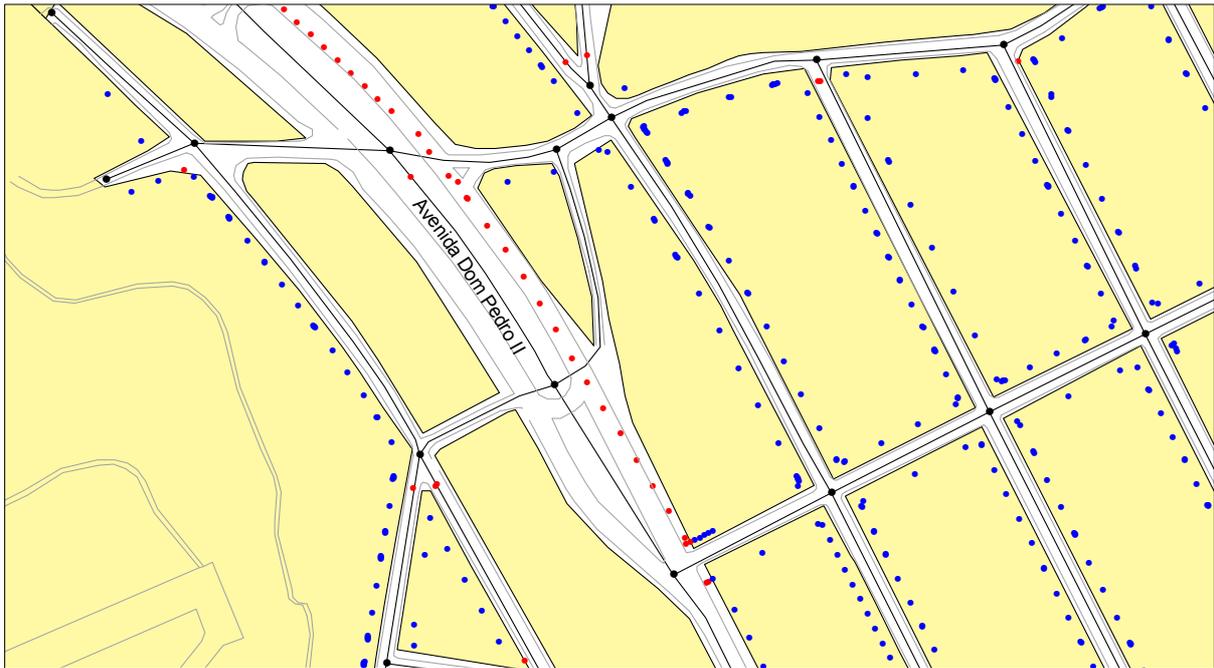


Figura 3.4.5 – Erro de posicionamento devido ao fator de afastamento inadequado. ⁷³

⁷³ Fonte: Imagem capturada da tela do mapa produzido pelo software ArcGis 8.3.

3.4.3 – Observação do Tratamento do Problema da Numeração Irregular

Conforme se esperava, a metodologia proposta proporcionou a correta geocodificação de trechos de logradouros contendo numeração irregular. O exemplo abaixo (figura 3.4.4) mostra a face 150364 da Rua Avelino dos Santos contendo numeração irregular. Os pontos em vermelho correspondem às posições geocodificadas dos endereços com a respectiva numeração de imóvel. No modelo de geocodificação habitual, o número 71, por exemplo, seria posicionado entre os números 65 e 161, contrariamente à realidade em campo. Este exemplo também mostra que haveria uma sobreposição de numeração máxima e mínima nos trechos da face, o que não permitiria uma correspondência perfeita dos endereços nestes trechos pela metodologia habitual.

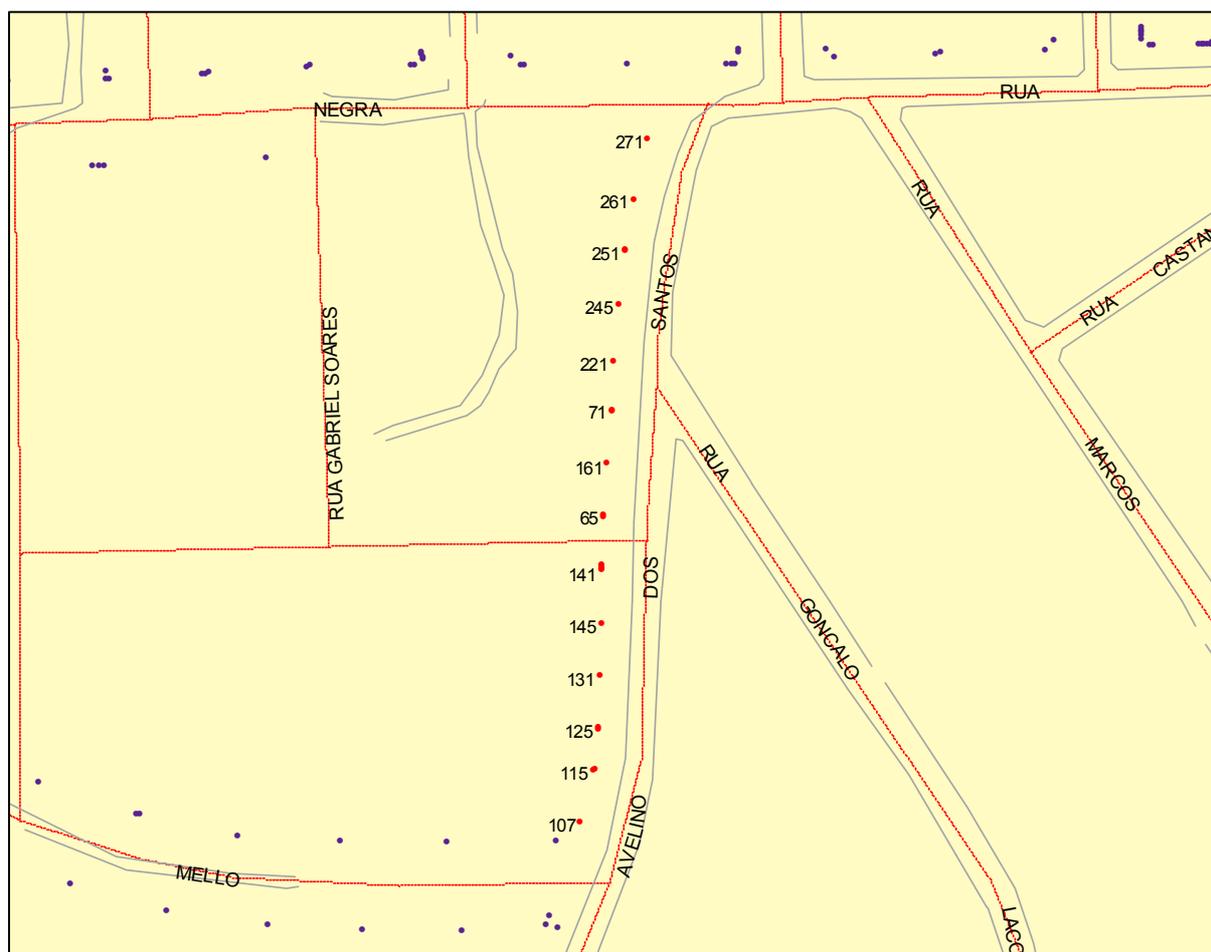


Figura 3.4.6 – Resultado da geocodificação da face 150364 contendo numeração irregular.⁷⁴

⁷⁴ Fonte: Imagem capturada da tela do mapa produzido pelo software ArcGis 8.3.

Capítulo IV – Conclusão

Os objetivos propostos nesta dissertação foram alcançados pelo desenvolvimento do sistema SIURB que se baseou na necessidade de produção de uma base de dados espaciais de clientes da COPASA. A partir de sua implantação e à medida que forem sendo cadastradas as rotas de leitura de consumo e entrega de faturas, esta base de dados estará sendo construída com a qualidade desejada para os propósitos de utilização previstos.

Neste trabalho, o objetivo não foi a realização de um estudo de caso específico utilizando GIS. O que se buscou foi o desenvolvimento de um Sistema de Informação que, estando em operação, poderá fazer parte do dia-a-dia de uma empresa de saneamento, inserindo-se nos seus processos rotineiros, produzindo resultados permanentes que serão de muita valia na análise de informações e na tomada de decisão que demandam dados em uma perspectiva espacial. O mesmo propiciará as bases para se alcançarem objetivos mais audaciosos e de retorno mais imediato, como a redução das perdas de água tratada, objetivo permanentemente buscado pelas empresas de saneamento em todo o mundo.

A proposta metodológica, aqui apresentada e utilizada, trouxe organização, padronização e, pelo aproveitamento de uma metodologia já utilizada internamente nos processos comerciais da empresa, produziu resultados melhores do que os atingidos pela metodologia habitual de geocodificação de endereços, além de possuir uma característica muito importante, ou seja, garantir a atualização constante da base de dados espacial produzida.

A análise dos resultados alcançados pela observação da taxa de geocodificação, do posicionamento dos pontos representativos dos endereços dentro dos limites das quadras e do tratamento dado ao problema da numeração irregular, mostra que a metodologia adotada neste trabalho trouxe uma contribuição positiva ao problema da geocodificação de endereços.

A correção dos erros de entrada de dados e a atualização dos trechos de logradouros com a realidade de campo nas áreas desatualizadas possibilitarão a geocodificação de 100% dos clientes da COPASA. O SIURB provê relatórios de consistência que facilitam esta correção.

A solução encontrada trouxe padronização e simplicidade aos processos envolvidos. A idéia de se trabalhar com a metodologia de leitura de consumo e entrega de faturas da COPASA não só permitiu a geocodificação dos endereços dos seus clientes, como a produção de mapas de rotas informatizados em substituição aos croquis em papel extremamente incorretos do ponto de vista cartográfico. Além disso, o SIURB trouxe uma integração com o sistema comercial, o que permitiu manter o cadastro de clientes georreferenciado no mesmo nível de atualização do cadastro de clientes do sistema comercial. Este é o caso, porque toda e qualquer alteração no cadastro de clientes do sistema comercial, que já possui um processo consolidado de manutenção, é automaticamente tratada também pelo SIURB na base de endereços de clientes georreferenciados.

A taxa de geocodificação de 97,4% é bastante significativa, principalmente se comparada ao processo habitual. Entretanto, esta taxa só foi alcançada porque o SIURB produziu relatórios de consistência que apontaram os erros encontrados nas bases de dados de referência. Tal como houve uma correção nesta base para se alcançar os resultados observados, as rejeições observadas no processo habitual poderiam ser trabalhadas e corrigidas para se alcançar melhores resultados. Entretanto, o processo habitual iria demandar não só correções na nomenclatura dos logradouros, mas também nas numerações máximas e mínimas nos trechos dos logradouros, além da inclusão dos trechos não encontrados da base de referência. Neste caso, a única fonte de referência seria a visitação em campo, exigindo também um tratamento individual exaustivo para os endereços rejeitados.

No caso da metodologia adotada, parte-se da premissa de que o cadastro de clientes da COPASA é um espelho bastante correto da realidade de campo, uma vez que toda e qualquer irregularidade do endereço é tratada de imediato, dada a necessidade de inclusão do cliente no processo de faturamento da empresa.

Já os resultados alcançados no posicionamento dos endereços podem ser melhorados, caso se desenvolva melhores algoritmos para a determinação dos fatores de afastamento e encurtamento. A identificação das situações de erro, conforme descritas no item 3.4.2, é um indicativo para se realizar melhorias futuras neste algoritmo. Entretanto, estes erros de posicionamento não comprometem a utilidade do arquivo de endereços georreferenciados obtido, uma vez que não inviabilizam as diversas análises espaciais que podem ser realizadas para a maioria das necessidades de informação, algumas delas identificadas na introdução deste trabalho.

Por último, o tratamento do problema da numeração irregular é uma contribuição interessante

desta metodologia, posto que este problema está presente em cerca de 5% dos logradouros em Belo Horizonte. A metodologia proposta não só corrige este problema, mas principalmente possibilita que os endereços nestes logradouros possam ser geocodificados, o que seria impossível pela metodologia habitual.

Há que se destacar, também, que o trabalho de atribuição dos códigos de faces nos trechos de logradouro possibilita, além da construção de uma base espacial de clientes, a produção de mapas de rotas de leitura em substituição aos arcaicos e imprecisos croquis. Esta possibilidade, embora já tenha sido observada antes do início dos trabalhos, contribuiu para justificar a utilização da metodologia proposta.

Embora não se tenha eliminado todos os problemas que o processo de geocodificação possui e que dependem, em muito, da qualidade dos dados, os resultados alcançados foram plenamente satisfatórios. Entretanto, os resultados ainda podem ser melhorados pelo aperfeiçoamento dos algoritmos nos ajustes dos fatores de afastamento e encurtamento, com vistas a produzir um melhor posicionamento espacial dos endereços. Neste caso, também, uma melhor definição das rotas de leitura pode reduzir as imprecisões de posicionamento verificadas. Todavia, o maior desafio continua sendo a desatualização da base de dados de referência que, dependendo de sua qualidade, pode influir muito negativamente na taxa de geocodificação, comprometendo a qualidade da base espacial de endereços produzida. Especial atenção, portanto, deve ser dispensada na aquisição ou criação de uma base de referência que tenha boa qualidade.

Apesar dos resultados alcançados atenderem aos requisitos de qualidade desejados pela COPASA, uma avaliação mais precisa poderia ser alcançada pela comparação geométrica dos endereços geocodificados com a base de endereços da prefeitura de Belo Horizonte. A base da prefeitura está georreferenciada pela coordenada de porta do imóvel, o que representa uma coordenada mais acurada do que a obtida pela interpolação do processo de geocodificação de endereços. Embora a base de endereços da prefeitura não esteja totalmente compatível com os endereços dos clientes da COPASA, essa comparação poderia fornecer uma medida mais objetiva do resultado do processo de geocodificação proposto. Embora fuja aos objetivos inicialmente estabelecidos nessa dissertação, essa análise poderia ser realizada em uma eventual continuação ou aprofundamento dessa pesquisa.

Por fim, acredita-se este trabalho não se aplique somente à Companhia de Saneamento de

Minas Gerais, mas que pode ser útil a qualquer empresa que possua uma metodologia de definição de rotas de visitaç o em campo semelhante   adotada pela COPASA e que queira representar seus clientes por suas coordenadas geogr ficas. Tamb m, novas perspectivas se abrem para a melhoria cont nua dos servi os prestados pelas empresas de servi o p blico. A cria o de uma base de dados georreferenciada de endere os com a qualidade que este trabalho prop e pode, com certeza, ser de grande valia para os diversos  rg os e empresas da administra o p blica, quer seja para o cruzamento de dados, quer seja para o simples, mas essencial conhecimento de onde se encontra o cidad o.

Referências Bibliográficas

- AMORIM FILHO, O. B., “A Geografia e as Qualidades Necessárias ao Geógrafo”, PUF, 1971. Traduzido de MEYNIER, A: “Guide de l’Étudiant em Géographie, Paris.
- ARONOFF, Stan, “Geographic Information Systems: A Management Perspective”, 3rd Edition, 1993, WDL Publications.
- BARTELS Richard H., John C. BEATTY and Brian A. BARSKY, “An Introduction To Splines for Use in Computer Graphics and Geometric Modeling”, Morgan Kaufmann Publishers, 1987.
- BRESLIN, Pat et all, “Getting to Know ArcView GIS”, ESRI. 1996.
- CANÇADO, Regina Darck. “Determinação de redes atingidas na interrupção do abastecimento de água: desenvolvimento de aplicativo computacional, utilizando GIS e grafos”. 2000. 126f. Dissertação (Mestrado em Tratamento da Informação Espacial) - Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte.
- CLARKE, K. C., “Getting Started with Geographic Information Systems”, Prentice Hall, 3rd. ed., 2000.
- COOKE, D.F. “Topology and TIGER: The Census Bureau's contribution.” In: T.W. Foresman, ed., *The History of Geographic Information Systems: Perspectives from the Pioneers*. Upper Saddle River, NJ: Prentice Hall, 1998.
- DRUMMOND, William J., “Address matching”, *Journal of the American Planning Association*; Spring95, Vol. 61 Issue 2, p240, 11p.
- ESRI, “Understanding GIS: The Arc/Info Method”, Redlands, CA: ESRI, 1990, pag. 1.2
- FOOTE, K.E., LYNCH, M. “GIS as an Integrating Technology: Contexts, Concepts, and Definitions”, Department of Geography, University of Texas at Austin, 1997. Disponível em <http://www.colorado.edu/geography/gcraft/notes/intro/intro_f.htm>. Acesso em 02 maio 2005.
- FOTHERINGHAM, S. & WEGENER, M.,”Spatial models and GIS: new potential and new models”, Ed. Taylor & Francis Group, 2000.
- FUSP: “Metodologia de Análise de Cadastros de Concessionárias”, São Paulo, Convênio ANEEL/FUSP, 2000.
- JARO, Matthew A., “Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida”, *Journal of the American Statistical Association*, 84,4:414-20, 1989.
- HARRIES, K., 1999, “Mapping Crime: Principles and Practice”, (Washington: US Department of Justice).

KNUTH, D. E., "The Art of Computer Programming", 1973, Volume 3: Sorting and Searching. Reading, MA: Addison-Wesley.

LONGLEY, Paul A., "Geographic Information Systems and Science", Ed. John Wiley & Sons, Ltd., 2001.

McMENAMIN, Stephen M.; PALMER, John F., "Análise Essencial de Sistemas"; tradutor Lars Gustav Erik Unonius; revisor técnico Fernando Manso, São Paulo, Ed. McGraw-Hill, 1991.

OLSON, Jack E., "Data Quality: The Accuracy Dimension", Morgan Kaufmann Publishers, San Francisco, 2003, pag. 24-27.

RATCLIFFE, Jerry H., "On the accuracy of TIGER-type geocoded address data in relation to cadastral and census areal units"; International Journal of Geographical Information Science, 2001, Vol. 15, no. 5, 475-485.

RICE, J. Richard, "Numerical Methods, Software, and Analysis", McGraw-Hill International Editions, 1983.

ROSA, E. S. & equipe, Saneamento Básico em Belo Horizonte: Trajetória em 100 anos, Rona Editora, FJP/CEHC, Belo Horizonte, 1996.

STAR, J. & ESTES, J., "Geographic Information Systems: An Introduction, Englewood Cliffs, NJ: Prentice Hall, 1990.

THOMAS O'REAGAN, R.; SAALFELD, A., "Geocoding Theory and Practice at the Bureau of the Census", Statistical Research Division Bureau of the Census, 1987.

TOMLIN, C. D., "Geographic Information Systems and Cartographic Modeling", Englewood Cliffs, NJ: Prentice Hall, 1990.

TOMLINSON, Roger F., "Thinking About GIS: geographic information system planning for managers" Redlands, CA: ESRI, 2003.

Anexos

- 1 – Modelo ambiental do SIURB
- 2 – Modelo comportamental do SIURB
- 3 – Descrição das bases de dados
- 4 – Diagrama físico do sistema
- 5 – Código AML do programa de geocodificação dos clientes
- 6 – Números e indicadores da COPASA

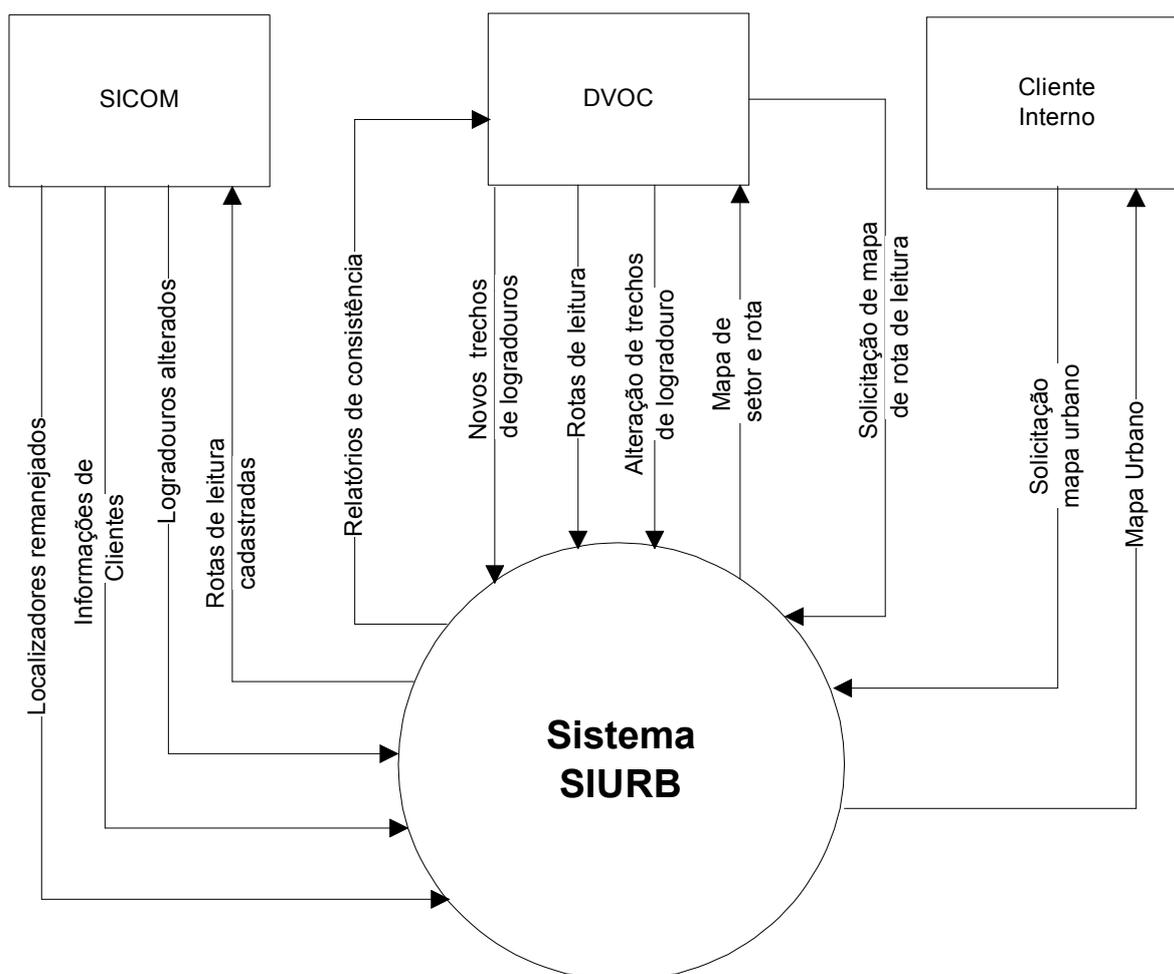
Anexo 1 – Modelo Ambiental do SIURB

- **Declaração dos objetivos do sistema**

O SIRUB tem por objetivo geocodificar os endereços dos clientes da COPASA a partir do modelo de representação espacial das rotas de leitura de consumo e entrega de faturas do sistema de faturamento da empresa. Busca também tornar disponíveis as seguintes informações geográficas urbanas por ele mantidas:

- municípios;
- bairros;
- logradouros;
- trechos de logradouros;
- clientes.

- **Diagrama de Contexto**



- **Descrição do Contexto**

O Sistema SIURB será implantado, inicialmente, na cidade de Belo Horizonte. Esta abrangência foi definida em função da cidade já possuir o cadastro de trechos de logradouros digitalizado no ambiente de geoprocessamento. A partir do momento em que outras cidades operadas pela COPASA possuírem o cadastro de trechos de logradouros em meio digital, o SIURB poderá ser adaptado e expandido para as mesmas.

Em Belo Horizonte, a divisão de operações comerciais (DVOC) é a área responsável pela criação e manutenção das rotas de leitura da COPASA. O sistema comercial SICOM é o ambiente em são processadas as alterações no cadastro de logradouros, bairros e no cadastro de clientes. Através de uma integração entre os sistemas, estas alterações são automaticamente representadas no SIURB que coloca à disposição dos clientes internos, na forma de mapas e atributos associados, toda a base de dados georreferenciada construída.

Desta forma, o contexto previsto para o sistema é a cidade de Belo Horizonte e as rotas de leitura inseridas para a geocodificação dos endereços dos clientes COPASA são de responsabilidade da DVOC.

- **Matriz de Eventos**

Matriz de Eventos

Nome do Sistema: Sistema de Informações Urbanas	Código do Sistema: SIURB
--	---------------------------------

Nº	EVENTO	ESTÍMULO	ORIGEM	AÇÃO	RESPOSTA	DESTINO	TIPO
1	DVOC cadastra rotas de leitura	Rotas a serem cadastradas	DVOC	Cadastrar rotas de leitura	Trechos de logradouros com as rotas de leitura cadastradas		E
2	DVOC atualiza trechos de logradouros	Trechos de logradouros a serem atualizados	DVOC	Atualizar trechos de logradouros	Trechos de logradouros atualizados		E
3	Momento de geocodificar clientes			Geocodificar clientes	Clientes geocodificados e relatórios de consistência	DVOC	T
4	Momento de atualizar localizadores remanejados	Localizadores remanejados no SICOM	SICOM	Atualizar localizadores remanejados	Localizadores de clientes e de trechos de logradouros alterados		T
5	Momento de atualizar logradouros	Logradouros atualizados no SICOM	SICOM	Atualizar descrição de logradouros	Trechos de logradouros com a descrição de logradouro alterada		T
6	DVOC solicita Mapa de rota de leitura	Solicitação de Mapa de rota de leitura	DVOC	Emitir Mapa de rota de leitura	Mapa de rota de leitura impresso	DVOC	E
7	Cliente interno solicita Mapa Urbano	Solicitação de Mapa Urbano	Cliente interno	Emitir Mapa Urbano	Mapa Urbano impresso	Cliente interno	E

Tipo: E - Externo T - Temporal C - Condicional

Anexo 2 – Modelo Comportamental

Diagrama de Entidades e Relacionamentos

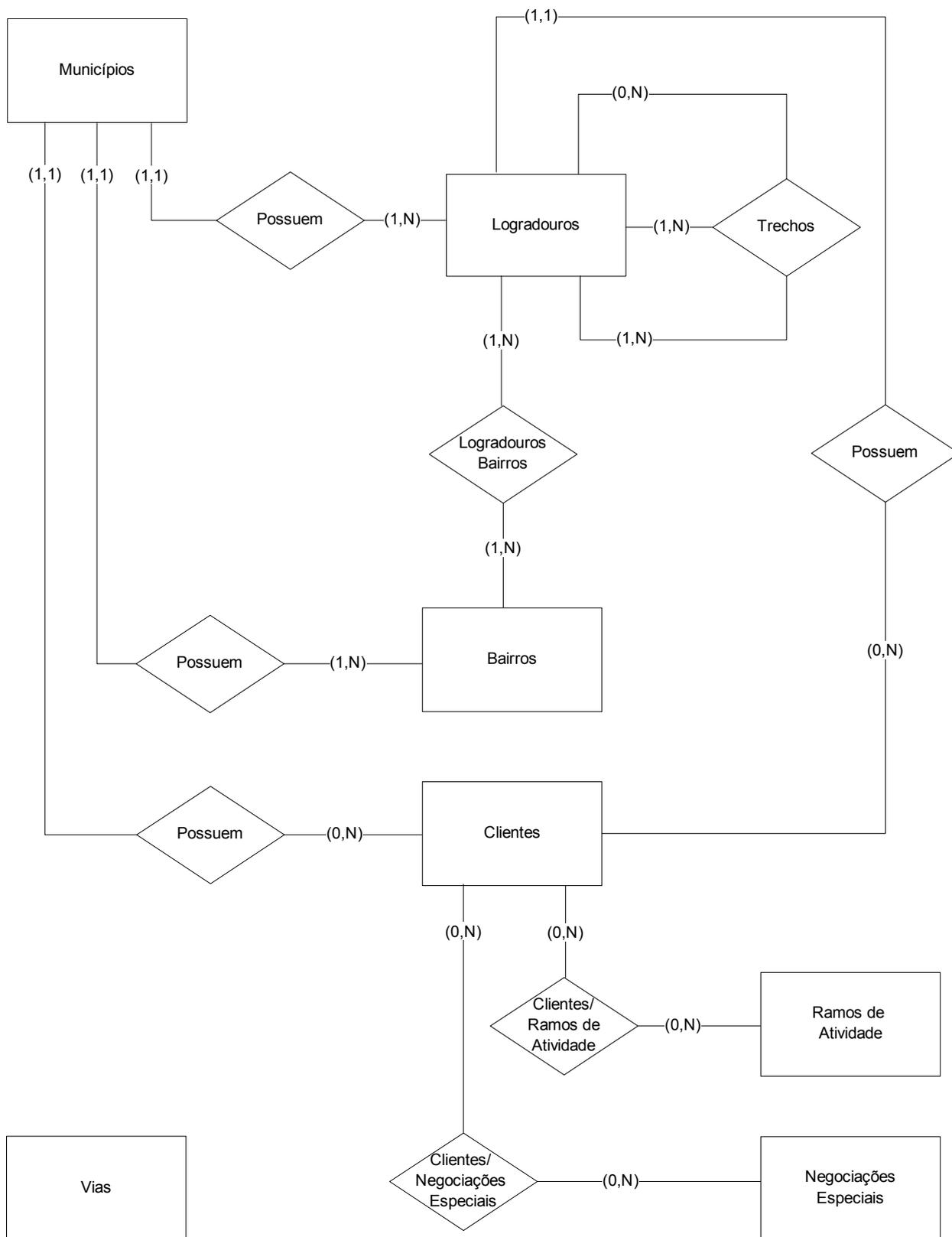
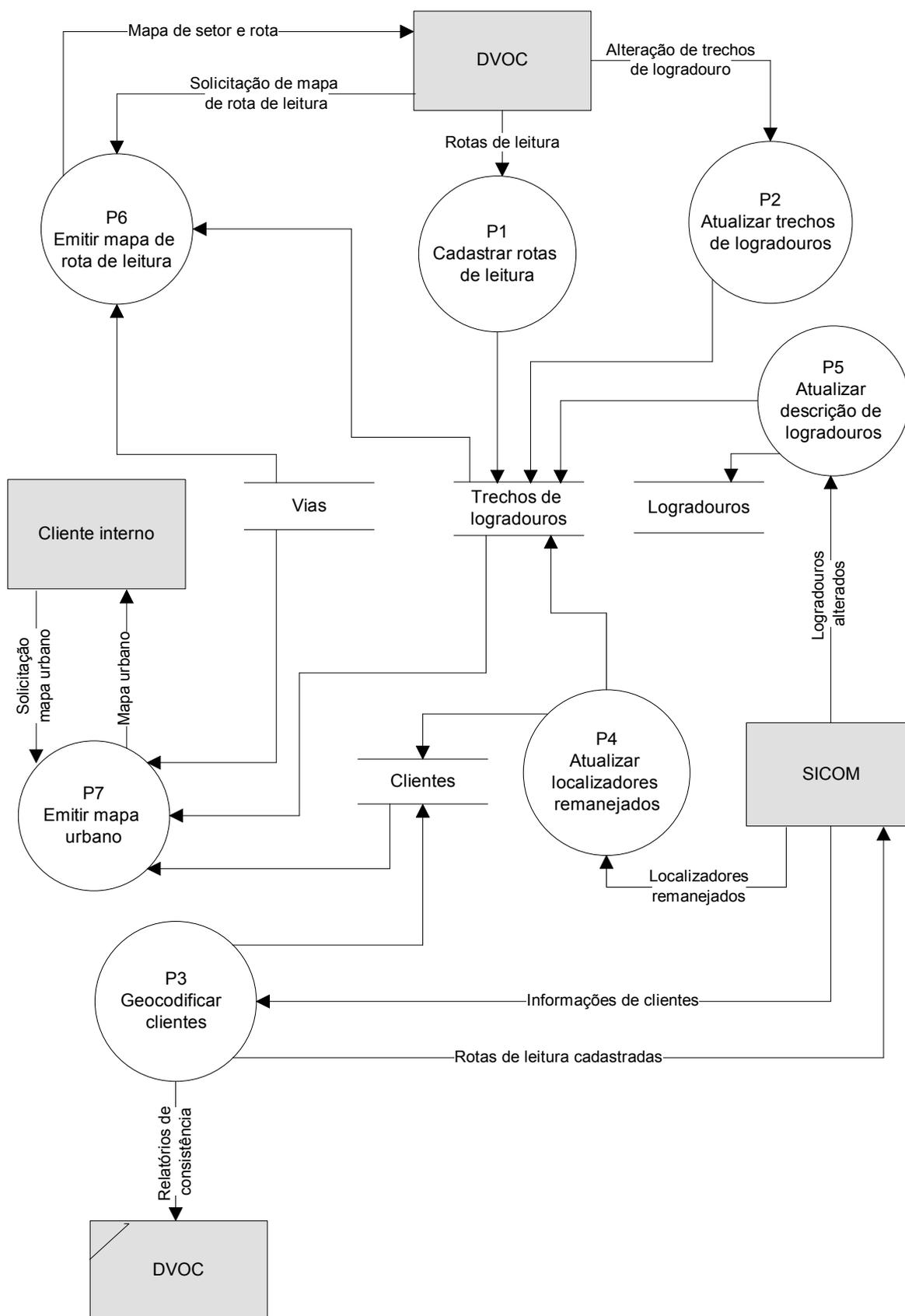


Diagrama de Fluxo de Dados Preliminar



Anexo 3 – Descrição das Bases de Dados

DESCRIÇÃO DO LAYER (COVERAGE)		
NOME: MUNICÍPIOS	CHAVE PRIMÁRIA: CD-MUNICÍPIO	
DESCRIÇÃO: Arquivo contendo todos os municípios de Minas Gerais	SIMBOLO E COR: Linha de contorno na cor branca	
FEATURES/DAT: MUNICÍPIOS.PAT MUNICÍPIOS.AAT MUNICÍPIOS.NAT	TOPONÍMIA: Nome do município.	
ATRIBUTOS ASSOCIADOS		
NOME	DESCRIÇÃO	TAMANHO E FORMATO
*** MUNICÍPIOS.PAT ***		
AREA	Área do polígono.	18,5F
PERIMETER	Perímetro.	18,5F
MUNICÍPIOS#	Atributo interno do software ArcInfo.	4/5 B
MUNICÍPIOS-ID	Atributo interno do software ArcInfo.	4/5 B
CD-MUNICÍPIO	Código da localidade conforme codificação utilizada no SICOM.	9 I
NO-MUNICÍPIO	Nome do município.	30 C

DESCRIÇÃO DO LAYER (COVERAGE)		
NOME: BAIRROS	CHAVE PRIMÁRIA: CD-MUNICÍPIO+CD-BAIRRO-COPASA	
DESCRIÇÃO: Arquivo contendo todos os bairros de Belo Horizonte.	SIMBOLO E COR: Linha pontilhada na cor branca	
FEATURES/DAT: BAIRROS.PAT BAIRROS.AAT BAIRROS.NAT	TOPONÍMIA: Nome do bairro.	
ATRIBUTOS ASSOCIADOS		
NOME	DESCRIÇÃO	TAMANHO E FORMATO
*** BAIRROS.PAT ***		
AREA	Área do polígono.	18,5F
PERIMETER	Perímetro.	18,5F
BAIRROS#	Atributo interno do software ArcInfo.	4/5 B
BAIRROS-ID	Atributo interno do software ArcInfo.	4/5 B
CD-MUNICÍPIO	Código da localidade conforme codificação utilizada no SICOM.	9 I
CD-BAIRRO-COPASA	Número seqüencial do bairro por localidade. A codificação será a mesma adotada pelo SICOM.	6 I
NO-BAIRRO-COPASA	Nome do bairro segundo a nomenclatura da COPASA.	30 C
NO-APELIDO	Apelido do bairro.	30 C

ATRIBUTOS ASSOCIADOS		
NOME	DESCRIÇÃO	TAMANHO E FORMATO
CD-BAIRRO-PREF	Código do bairro conforme padrão da prefeitura	6 I
NO-BAIRRO-PREF	Nome do bairro segundo a nomenclatura da Prefeitura.	30 C
DT-ULT-ALTERACAO	Data última atualização.	8 I
CD-SIMBOLO	Código do símbolo.	3 I
CD-BAIRRO	Código do bairro conforme padrão da prefeitura.	4 I
CD-BAIRRO-SICOM	Código do bairro conforme SICOM.	6 I

DESCRIÇÃO DO ARQUIVO (INFOFILE)		
NOME: BAIRROS-LOGRADS	CHAVE PRIMÁRIA: CD-MUNICIPIO+CD-BAIRRO + CD-LOGRADOURO	
DESCRIÇÃO: Arquivo resultante da interpolação dos arquivos de Trechos e Bairros. Contendo todos os bairros percorridos pelos logradouros de Belo Horizonte.	SIMBOLO E COR: Não se aplica.	
FEATURES/DAT: -----xxx-----	TOPONÍMIA: Não se aplica.	
ATRIBUTOS ASSOCIADOS		
NOME	DESCRIÇÃO	TAMANHO E FORMATO
CD-MUNICIPIO	Código da localidade conforme codificação utilizada no SICOM.	9 I
CD-BAIRRO	Número seqüencial do bairro por localidade. A codificação será a mesma adotada pelo SICOM.	6 I
CD-LOGRADOURO	Número seqüencial do logradouro na localidade. A codificação é a mesma adotada pelo SICOM.	6 I
DT-ULT-ALTERACAO	Data última atualização.	8 I

DESCRIÇÃO DO ARQUIVO (INFOFILE)		
NOME: TABELA--NEGOCIACOES-ESPECIAIS	CHAVE PRIMÁRIA: CD-TIPO-NEGOCIACAO	
DESCRIÇÃO: Descreve o tipo de negociação especial realizada com o cliente.	SIMBOLO E COR: Não se aplica.	
FEATURES/DAT: -----xxx-----	TOPONÍMIA: Não se aplica.	
ATRIBUTOS ASSOCIADOS		
NOME	DESCRIÇÃO	TAMANHO E FORMATO
CD-TIPO-NEGOCIACAO	Tipo de negociação especial realizada com o cliente.	2 C
NO-NEGOCIACAO-ESPECIAL	Descrição da negociação especial.	30 C

DESCRIÇÃO DO LAYER (COVERAGE)		
NOME: CLIENTES	CHAVE PRIMÁRIA: NU-MATRICULA	
DESCRIÇÃO: Identificação dos clientes da empresa.	SÍMBOLO E COR: 550 – (casinha azul).	
FEATURES/DAT: CLIENTES.PAT	TOPONÍMIA: Número do imóvel.	
ATRIBUTOS ASSOCIADOS		
NOME	DESCRIÇÃO	TAMANHO E FORMATO
*** CLIENTES.PAT ***		
AREA	Área do polígono.	18,5F
PERIMETER	Perímetro.	18,5F
CLIENTES#	Atributo interno do software ArcInfo.	4/5 B
CLIENTES-ID	Atributo interno do software ArcInfo.	4/5 B
NU-MATRICULA	Número seqüencial único que identifica o cliente. É a matrícula do cliente.	11 I
CD-MUNICIPIO	Código do município.	9 I
CD-SRFSEQ	O localizador é composto pela numeração do Setor-rota-face-sequencia por localidade, com o seguinte formato: Setor - 2 posições Rota - 2 posições Face - 2 posições Seqüência - 3 posições	9 C
CD-SRFACE	Redefinição do campo CD-SRFSEQ de suas 6 primeiras posições.	6 C
CD-SEQ	Redefinição do campo CD-SRFSEQ de suas três últimas posições.	3 C
CD-LOGRADOURO	Identificação do logradouro onde o imóvel se localiza, conforme codificação utilizada no SICOM.	6 I
NU-TRECHO	Número de identificação do trecho onde o imóvel está localizado	10 I
ID-LADO	Identificação do lado do trechos onde o imóvel está localizado. D - lado direito do trecho E - lado esquerdo do trecho	1 C
NU-IMOVEL	Número do imóvel	5 I
CD-TIPO-COMPL	Tipo de complemento do endereço do imóvel, conforme Tabela de tipo de complemento do SICOM.	2 C
DE-COMPLEMENTO	Complemento do endereço do imóvel	12 C
CD-USUARIO	Identificação de usuário que executou a última alteração.	8 C
DT-ULT-ALTERACAO	Data da última alteração.	8 I

DESCRIÇÃO DO ARQUIVO (INFOFILE)		
NOME: LOGRADOUROS	CHAVE PRIMÁRIA: CD-LOGRADOURO	
DESCRIÇÃO: Tabela contendo todos os logradouros de Belo Horizonte.	SÍMBOLO E COR: Não se aplica.	
FEATURES/DAT: -----xxx-----	TOPONÍMIA: Não se aplica.	
ATRIBUTOS ASSOCIADOS		
NOME	DESCRIÇÃO	TAMANHO E FORMATO
CD-MUNICIPIO	Código da localidade conforme codificação utilizada no SICOM.	9 I
CD-LOGRADOURO	Número seqüencial do logradouro na localidade. A codificação é a mesma adotada pelo SICOM.	6 I
TP-LOGRADOURO	Tipo do logradouro conforme Tabela Tipo-logradouro do SICOM.	2 C
NO-LOGRADOURO	Nome logradouro segundo nomenclatura da COPASA. Corresponde ao nome popular do logradouro para a Prefeitura	40 C
NO-APELIDO-1	Apelido logradouro. É outro nome pelo qual o logradouro é conhecido	40 C
NO-APELIDO-2	Apelido logradouro. É outro nome pelo qual o logradouro é conhecido.	40 C
CD-PREFEITURA	Código do logradouro segundo a prefeitura.	7 I
DT-ULT-ALTERACAO	Data última atualização.	8 I
DT-INC-TRECHO	Data de inclusão do logradouro no arquivo de TRECHOS.	8 I

DESCRIÇÃO DO ARQUIVO (INFOFILE)		
NOME: SU-CONTROLE-EXECUCAO	CHAVE PRIMÁRIA: CD-ACESSO	
DESCRIÇÃO: Controle de execução de programas Batch do SIURB.	SÍMBOLO E COR: Não se aplica.	
FEATURES/DAT: -----xxx-----	TOPONÍMIA: Não se aplica.	
ATRIBUTOS ASSOCIADOS		
NOME	DESCRIÇÃO	TAMANHO E FORMATO
CD-ACESSO	Código de acesso de execução da fase.	22 C
DT-INICO-EXECUCAO	Data de início da execução da fase.	8 I
HR-INIC-EXECUCAO	Hora de início de execução da fase.	6 I
DT-FIM-EXECUCAO	Data do fim de execução da fase.	8 I
HR-FIM-EXECUCAO	Hora do fim de execução da fase.	6 I
DE-INF-LIVRE	Mensagem de execução.	40 C

DESCRIÇÃO DO LAYER (COVERAGE)		
NOME: TRECHOS	CHAVE PRIMÁRIA: NU-TRECHO	
DESCRIÇÃO: Linha de centro do logradouro definida pelo cruzamento de logradouros. A ser carregado no banco de dados Arcstorm.	SIMBOLO E COR: De acordo com o atributo CD-SIMBOLO.	
FEATURES/DAT: TRECHOS.AAT TRECHOS.NAT	TOPONÍMIA: Nome do logradouro.	
ATRIBUTOS ASSOCIADOS		
NOME	DESCRIÇÃO	TAMANHO E FORMATO
*** TRECHOS.AAT ***		
TRECHOS#	Atributo interno do software ArcInfo.	4/5 B
TRECHOS-ID	Atributo interno do software ArcInfo.	4/5 B
FNODE#	Nó de início.	
TNODE#	Nó de término.	
LENGTH	Comprimento do arco.	4/5 B
NU-TRECHO	Número único de identificação do trecho na COPASA.	10 I
CD-MUNICIPIO	Código da localidade conforme codificação utilizada no SICOM.	9 I
CD-LOGRADOURO	Código do logradouro Principal, conforme codificação utilizada no SICOM.	6 I
TP-LOGRADOURO	Tipo do logradouro conforme Tabela de Tipo de logradouros do SICOM.	2 C
NO-LOGRADOURO	Nome do logradouro conforme Tabela de logradouros do SICOM.	40 C
NU-MIN-DIR	Menor número de imóvel existente no trecho em referência, do lado direito do mesmo.	5 I
NU-MAX-DIR	Maior número de imóvel existente no trecho em referência, do lado direito do mesmo.	5 I
NU-MIN-ESQ	Menor número de imóvel existente no trecho em referência, do lado esquerdo do mesmo.	5 I
NU-MAX-ESQ	Maior número de imóvel existente no trecho em referência, do lado esquerdo do mesmo.	5 I
NUTRH	Número único de identificação do trecho pela Prefeitura.	6 I
NULOG	Número único de identificação do logradouro pela Prefeitura	4/ 6 B
TPLOG	Tipo do logradouro na PRODABEL.	3 C
NOLOG	Nome do logradouro na PRODABEL.	50 C
NUMINDRT	Menor número de imóvel existente no trecho em referência, do lado direito do mesmo.	4/ 5 B
NUMAXDRT	Maior número de imóvel existente no trecho em referência, do lado direito do mesmo.	4/ 5 B
NUMINESQ	Menor número de imóvel existente no trecho em referência, do lado esquerdo do mesmo.	4/ 5 B
NUMAXESQ	Maior número de imóvel existente no trecho em referência, do lado esquerdo do mesmo.	4/ 5 B
CD-SIMBOLO	Código do símbolo gráfico.	3 I

ATRIBUTOS ASSOCIADOS		
NOME	DESCRIÇÃO	TAMANHO E FORMATO
CD-SRFACE-DIR	Numeração da face da direita. Possui o seguinte formato: Ss rr ff- ss- setor rr- rota ff- face	6 C
CD-SRFACE-ESQ	Numeração da face da esquerda. Possui o seguinte formato: Ss rr ff- ss- setor rr- rota ff- face	6 C
FG-TR-IRREGULAR	Flag de identificação se trecho irregular. Valores- S ou N	1 C
FG-PROCESSA-DIR	S - setor-rota-face-direita marcado para processamento dos imóveis N - setor-rota-face-direita já processado	1 C
FG-PROCESSA-ESQ	S - setor-rota-face-esquerda marcado para processamento dos imóveis N - setor-rota-face-esquerda já processado	1 C
ID-LOG	Contém valor constante = "A". Usado para processamento da localização geográfica dos imóveis.	1 C
NU-AFASTAMENTO	Valor usado para processamento da localização geográfica dos imóveis. Indica a distância de posicionamento dos imóveis em relação à center-line.	2 I

DESCRIÇÃO DO LAYER (COVERAGE)		
NOME: VIAS	CHAVE PRIMÁRIA: VIAS-ID	
DESCRIÇÃO: Limite do leito da via pública.	SIMBOLO E COR: De acordo com o atributo CD-SIMBOLO	
FEATURES/DAT: VIAS.AAT VIAS.ANNO	TOPONÍMIA: Nome do logradouro.	
ATRIBUTOS ASSOCIADOS		
NOME	DESCRIÇÃO	TAMANHO E FORMATO
*** VIAS.AAT ***		
VIAS#	Identificação interna do software.	4/5 B
VIAS-ID	Identificação interna do software.	4/5 B
LENGTH	Comprimento do arco.	4/5 B
CD-SIMBOLO	Contém o código do símbolo para cada tipo de via: Via pavimentada com meio-fio; Via não pavimentada com meio-fio; Via pavimentada sem meio-fio; Via não pavimentada sem meio-fio; Rodovia; Acostamento; Linha de trem; Ponte; Viaduto; Outro.	3 I

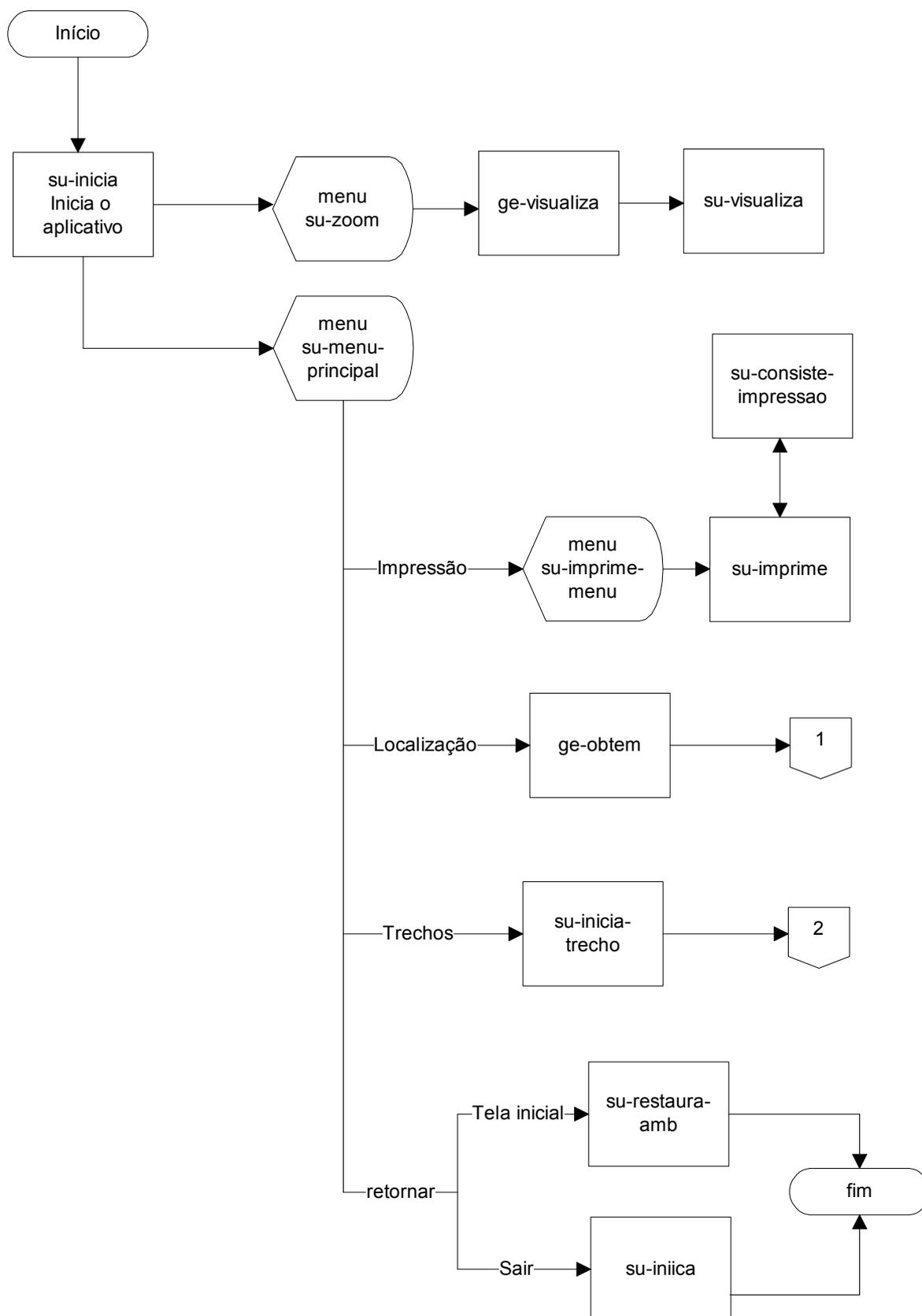
DESCRIÇÃO DO ARQUIVO (INFOFILE)		
NOME: CLIENTES-RAMOS	CHAVE PRIMÁRIA: NU-MATRICULA + CD-RAMO-ATIVIDADE	
DESCRIÇÃO: Associa o ramo de atividade do imóvel ao cliente.	SÍMBOLO E COR: Não se aplica.	
FEATURES/DAT: -----xxx-----	TOPONÍMIA: Não se aplica.	
ATRIBUTOS ASSOCIADOS		
NOME	DESCRIÇÃO	TAMANHO E FORMATO
NU-MATRICULA	Número seqüencial único que identifica o cliente.	11 I
CD-RAMO-ATIVIDADE	Código do ramo de atividade do imóvel.	4 I

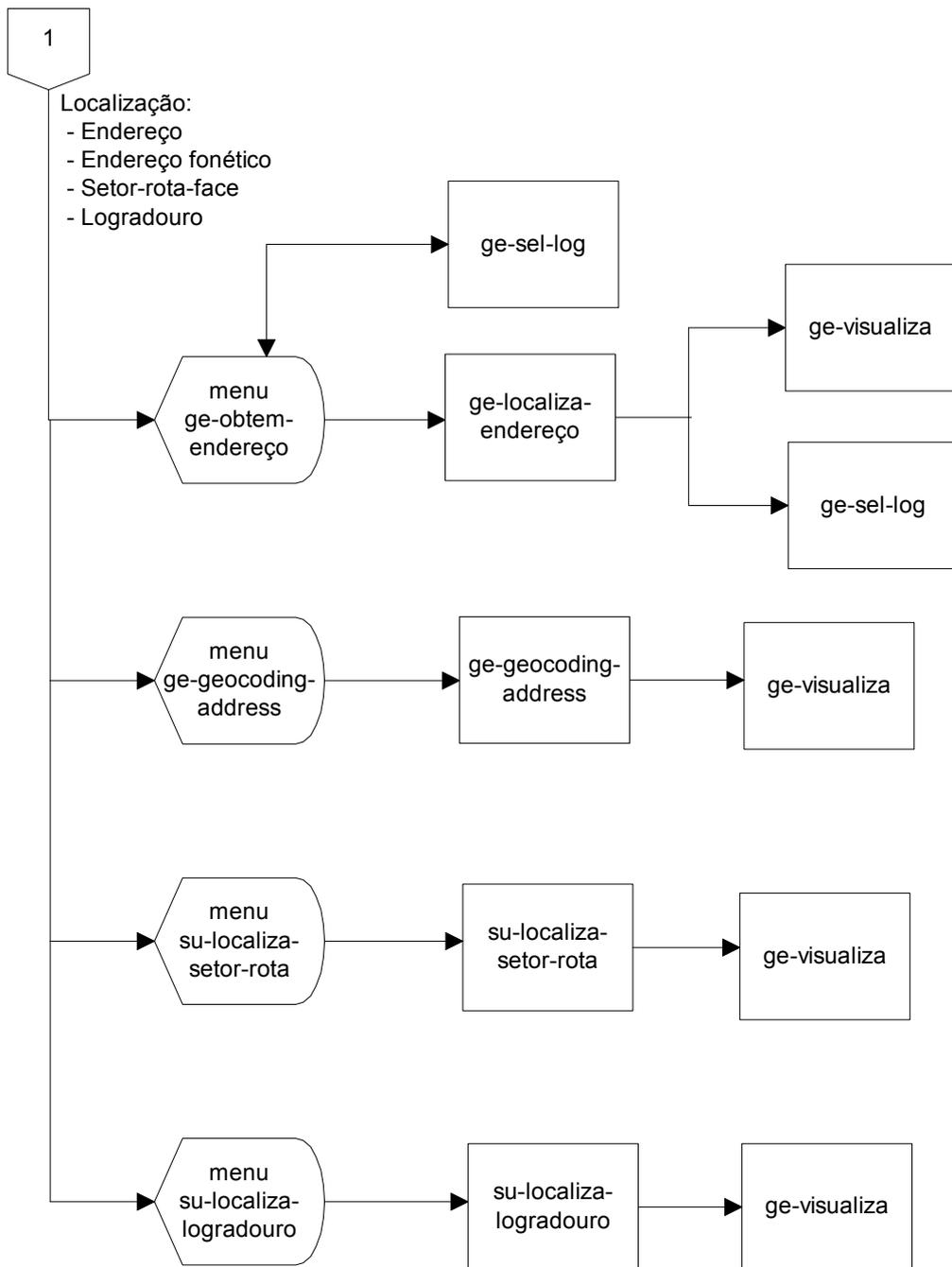
DESCRIÇÃO DO ARQUIVO (INFOFILE)		
NOME: CLIENTES-NEGOCIAÇÃO-ESPECIAL	CHAVE PRIMÁRIA: NU-MATRICULA + CD-TIPO-NEGOCIACAO	
DESCRIÇÃO: Associa o tipo de negociação especial realizada com o cliente.	SÍMBOLO E COR: Não se aplica.	
FEATURES/DAT: -----xxx-----	TOPONÍMIA: Não se aplica.	
ATRIBUTOS ASSOCIADOS		
NOME	DESCRIÇÃO	TAMANHO E FORMATO
NU-MATRICULA	Número seqüencial único que identifica o cliente.	11 I
CD-TIPO-NEGOCIACAO	Tipo de negociação especial realizada com o cliente.	2 C

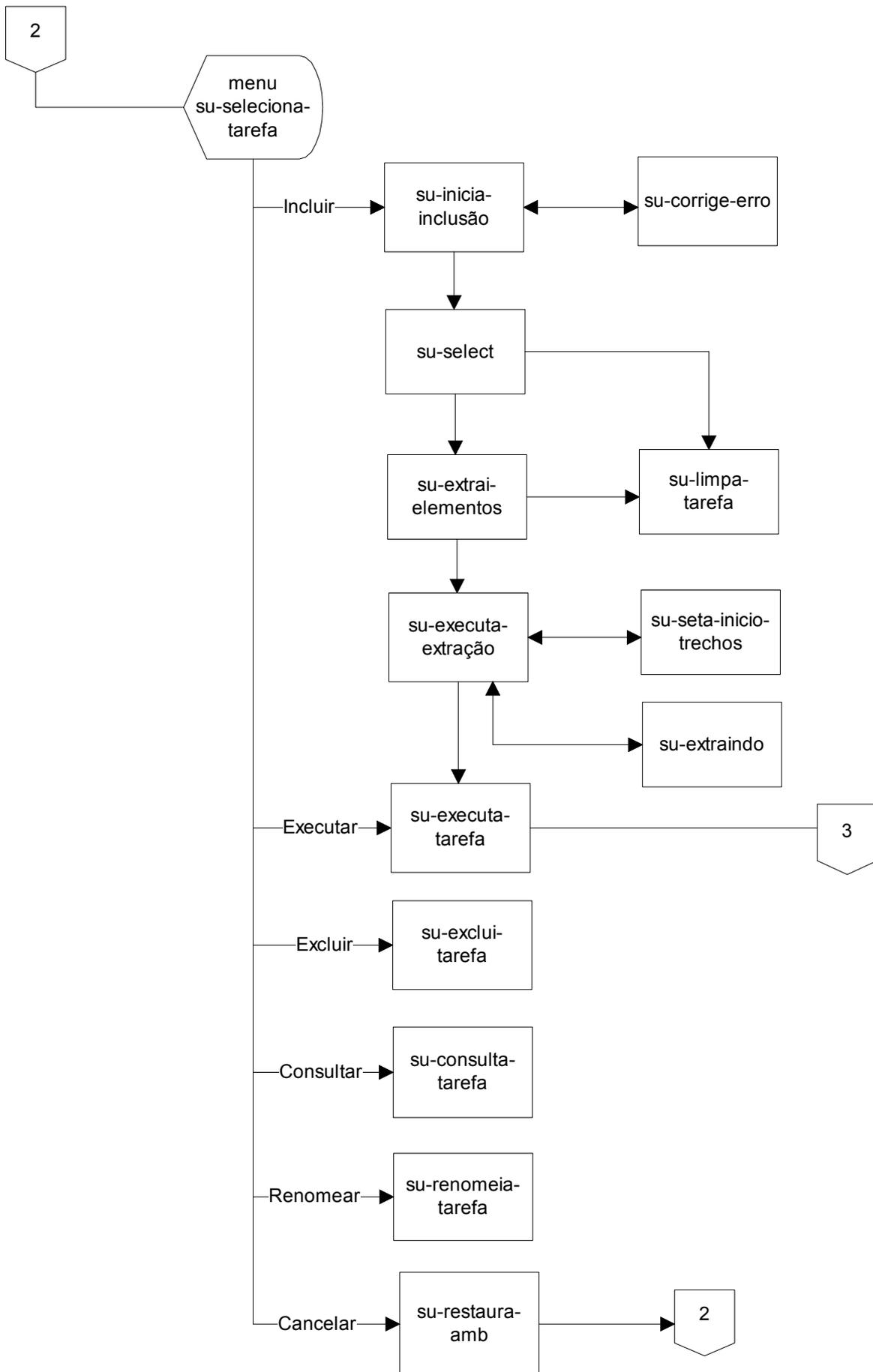
DESCRIÇÃO DO ARQUIVO (INFOFILE)		
NOME: TABELA-RAMOS-ATIVIDADE	CHAVE PRIMÁRIA: CD-RAMO-ATIVIDADE	
DESCRIÇÃO: Descreve o ramo de atividade do imóvel.	SÍMBOLO E COR: Não se aplica.	
FEATURES/DAT: -----xxx-----	TOPONÍMIA: Não se aplica.	
ATRIBUTOS ASSOCIADOS		
NOME	DESCRIÇÃO	TAMANHO E FORMATO
CD-RAMO-ATIVIDADE	Código do ramo de atividade do imóvel.	4 I
NO-RESUMIDO-RAMO-ATIVIDADE	Nome resumido do ramo de atividade.	15 C

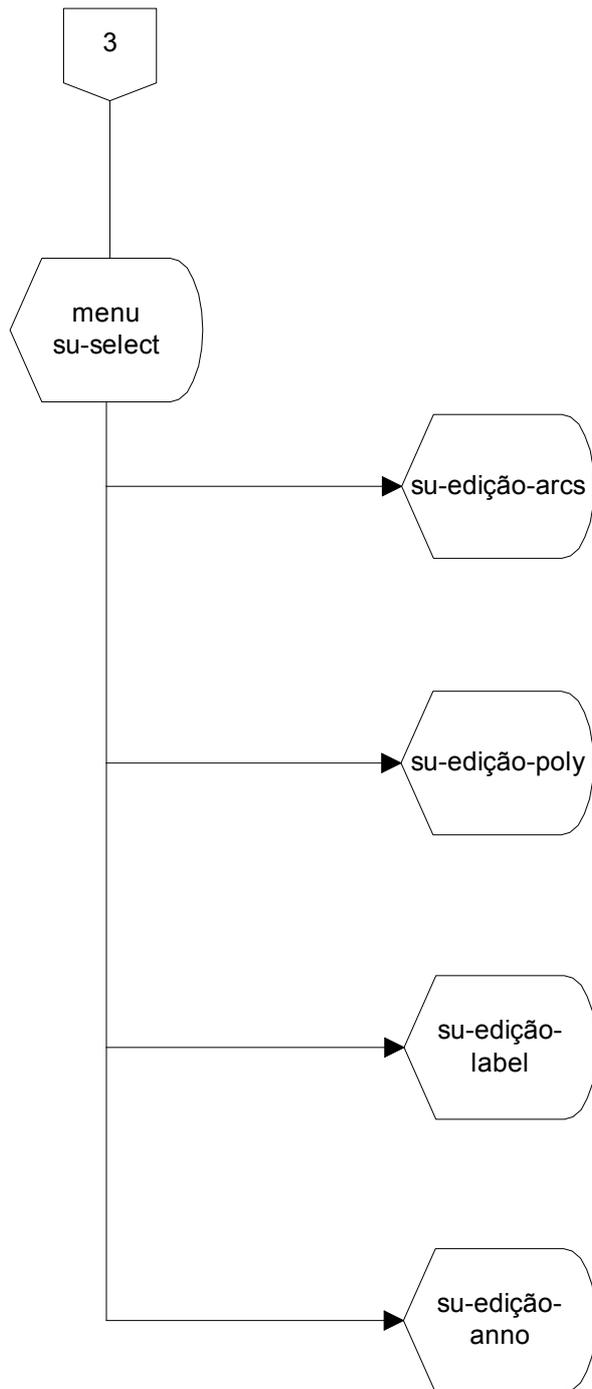
Anexo 4 - Diagrama físico do sistema

Processamento *online*



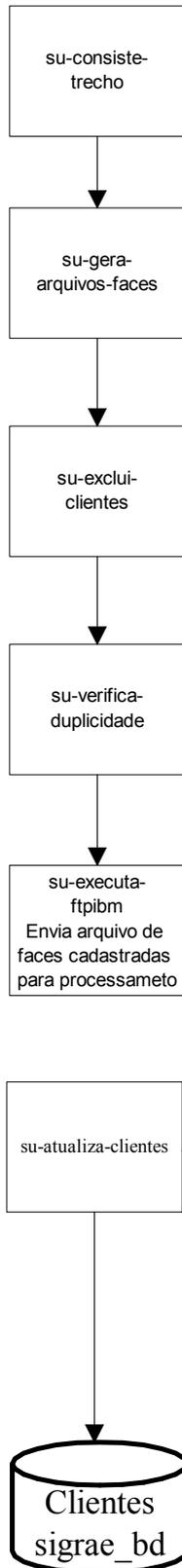
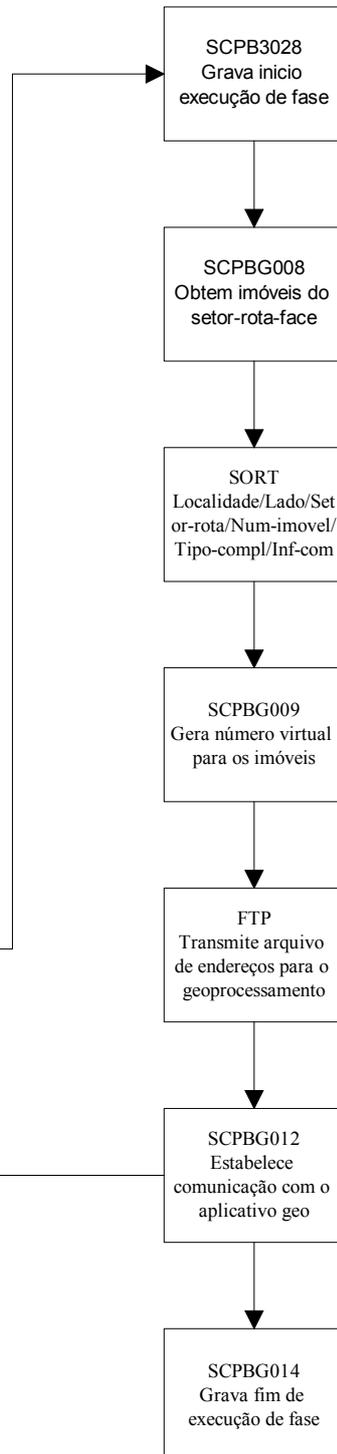






Processamento *batch*

Ambiente Geo

Ambiente Mainframe
FASE HPSCEG05

Anexo 5 - Código AML do programa de geocodificação dos clientes

```

/*-----
/* Procedimentos iniciais
/*-----

&args .fase .dtsub .hrsub .parametro-geral:REST

&severity &error &routine su-erro

&amlpath $SIURB_AP/fonte/aml/su
&watch $SIURB_WS/logs/su-atualiza-clientes-log &commands
&s pathlog $SIGRAE_WS/logs/atualiza-dvmo/erros

&if not [exists $SIURB_WS/xxsu%.dtsub%%.hrsub% -work] &then
    cw $SIURB_WS/xxsu%.dtsub%%.hrsub%

&workspace $SIURB_WS/xxsu%.dtsub%%.hrsub%

&if [show program] = ARCPLOT &then quit
&if [show program] = ARCEDIT &then quit

&date_format YMD/.
&s xxdata [calc [date -dfmt]]

&sys cp %diretorio%/home/sigrae-d/ftpibm/su-imovs
$SIURB_DA/tabela/wimoveis.%xxdata%
&sys cp %diretorio%/home/sigrae-d/ftpibm/su-imovs wimoveis.%xxdata%

&s txt1 [open clientes-del.txt openstatus -write]
&if [value openstatus] <> 0 &then &call su-erro
copyinfo $SIURB_DA/tabela/imoveis_copasa imoveis_copasa
copyinfo $SIURB_DA/tabela/srf_unicos srf_unicos

&data ARC TABLES
  SEL IMOVEIS_COPASA
  PURGE
  YES
  ADD FROM wimoveis.%xxdata%
  SORT MUNICIPIO-SRF
  Q
&end

/* Cria um infofile de trabalho com setor-rota-face unicos

&s totalsrf [listunique imoveis_copasa -info municipio-srf xxsr%.xxdata%]

&data ARC TABLES
  SEL SRF_UNICOS
  PURGE
  YES
  ADD FROM XXSRF.%xxdata%
  Quit
&end

arcplot
lib sigrae_bd.copasa
asconnect sigrae_bd

resel .trechos arc
asel .trechos arc
writesel sel_trechos .trechos arc
asexecute sigrae_bd setaction copyout copasa
asexecute sigrae_bd addlayer .trechos arc trechos
asexecute sigrae_bd setselfile sel_trechos

```

```

asexecute sigrae_bd execute
arc indexitem trechos.aat cd-srface-dir
arc indexitem trechos.aat cd-srface-esq
resel trechos arc
asel trechos arc
calc trechos arc nu-min-dir = 0
calc trechos arc nu-max-dir = 0
calc trechos arc nu-min-esq = 0
calc trechos arc nu-max-esq = 0

clearselect
resel .clientes point
writeselect sel_clientes

asexecute sigrae_bd setaction copyout copasa
asexecute sigrae_bd addlayer .clientes point clientes
asexecute sigrae_bd setselfile sel_clientes
asexecute sigrae_bd execute

&s min1 = nu-min-esq
&s max1 = nu-max-esq
&s min2 = nu-min-dir
&s max2 = nu-max-dir

resel srf_unicos info municipio-srf > ' '
writesel srf-sel srf_unicos info
&s qtde_srf_unicos [extract 1 [show select srf_unicos info]]

/*-----
/* Laço principal
/*-----
&do n = 1 &to [value qtde_srf_unicos]

  readsel srf-sel
  &s cd-municipio [show select srf_unicos info [value n] item cd-municipio]
  &s srface [show select srf_unicos info [value n] item srface]
  clearsel trechos arc
  resel trechos arc cd-srface-dir = [quote [value srface]]
  resel trechos arc cd-municipio = [value cd-municipio]
  &s arcos [extract 1 [show select trechos arc]]

  &if [value arcos] <> 0 &then /* Trata o setor-rota-face da direita
  &do
    &s parimpar = 2
    &call minmax
    &call clientes
  &end
  &else
  &do
    clearsel trechos arc
    resel trechos arc cd-srface-esq = [quote [value srface]]
    resel trechos arc cd-municipio = [value cd-municipio]
    &s arcos [extract 1 [show select trechos arc]]

    &if [value arcos] <> 0 &then /* Trata o setor-rota-face da esquerda
    &do
      &s parimpar = 1
      &call minmax
      &call clientes
    &end
    &else
    &do
      &sys echo Nao encontrou trecho para o setor-rota-face = [value srface]
      > $SIURB_WS/logs/xxsrf
    &end
  &end
&end /* Fim do laço principal

```

```

quit /* Sai do arcplot
/*-----
/* Procedimentos Finais
/*-----

/* Exclui campos indesejaveis

tables
dropitem clientes.pat xxsr# xxsr-id xxsr-side
quit

arcedit

/* Efetua o carga dos clientes novos

lib sigrae_bd.copasa
asconnect sigrae_bd

asexecute sigrae_bd setaction load copasa
asexecute sigrae_bd addlayer .clientes point clientes
asexecute sigrae_bd execute

/*

&if [exists $SIURB_WS/logs/su-atualiza-clientes.txt -file] &then
    &sys rm $SIURB_WS/logs/su-atualiza-clientes.txt
&s dt-hoje [date -vfull]
&s texto Pgm su-atualiza-clientes.aml(%.fase%) executado c/ sucesso em %dt-
hoje%
&s fecha [close -all]
&s a [open $SIURB_WS/logs/su-atualiza-clientes.txt opst -append]
&s w [write %a% [quote %texto%]]
&s fecha [close -all]

/* Executa programa de atualização de numeração de trechos

&workspace ..
&r $SIURB_AP/fonte/aml/su/su-gera-numeracoes-trecho
&workspace ..
&end

quit
quit
&end

&watch &off

&return /* Fim do Programa

/*-----
/* Sub-rotinas chamadas
/*-----

&routine minmax
/* Determina a numeração maxima e minima do trecho de logradouro

    &if [exists xxsr -coverage] &then arc kill xxsr all
    &if [exists xxsel -file] &then &sys rm xxsel
        /* Cria uma coverage temporaria do trecho do srf lido
writesel xxsel
arc reselect trechos xxsr arc xxsel

        /* Gera um arquivo info temporario com todos os clientes do srf
clearsel imoveis_copasa info
&if [exists xxinfo -info] &then arc killinfo xxinfo
resel imoveis_copasa info cd-srf = [quote [value srf]]

```

```

resel imoveis_copasa info cd-municipio = [value cd-municipio]
&s imoveis [extract 1 [show select imoveis_copasa info]]
infofile imoveis_copasa info xxinfo
&if [exists xxinfoaux -info] &then arc killinfo xxinfoaux
arc copyinfo xxinfo xxinfoaux

&if [value imoveis] >= [value arcos] &then
  &do
    &s parcela [truncate [calc [value imoveis] / [value arcos]]]
    &s resto [calc [value imoveis] - [value parcela] * [value arcos]]
    &s limite [value arcos]
  &end
&else
  &do
    &s parcela 1
    &s resto 0
    &s limite [value imoveis]
  &end

&s imin = 1
&s imax = [calc [value parcela] + [value resto]]

mapex arc xxsrvf
resel xxsrvf arc mapex
sort xxsrvf arc fnode#
clearsel xxinfoaux info
resel xxinfoaux info nu-virtual > 0
sort xxinfoaux info nu-virtual
&s numin [show select xxinfoaux info 1 item nu-virtual]
&s numax [show select xxinfoaux info [value imoveis] item nu-virtual]

/* atualiza a numeracao max e min do xxsrvf
&do i = 1 &to [value limite]
  &s nutrecho [show select xxsrvf arc [value i] item nu-trecho]
  &s numinimo [show select xxinfoaux info [value imin] item nu-virtual]
  &s numaximo [show select xxinfoaux info [value imax] item nu-virtual]
  clearsel xxinfo info
  resel xxinfo info nu-virtual <= [value numaximo]
  and nu-virtual >= [value numinimo]
  calc xxinfo info nu-trecho = [value nutrecho]
/*
  &s imin = [calc [value imax] + 1]
  &s imax = [calc [value imax] + [value parcela]]
&end

/* Faz o unsplit
&if [value arcos] > 1 &then
  &do
    quit
    arcedit
    ec xxsrvf arc
    sel all
    unsplit none
    save
    quit
    arcplot
  &end
mapex arc xxsrvf
resel xxsrvf arc mapex
calc xxsrvf arc [value min[value parimpar]] = [value numin]
calc xxsrvf arc [value max[value parimpar]] = [value numax]

/* Determina o squeeze factor
&s comprimento [show select xxsrvf arc 1 item length]
&s squeeze = [calc 2400 / [value comprimento] ]
&if [value squeeze] > 40 &then
&if [value squeeze] <= 55 &then &s squeeze = 40

```

```

&else
&if [value squeeze] <= 70 &then &s squeeze = 30
&else
&s squeeze = 20

clearsel xxsrfr arc

&return

/*-----
/* Geocodifica os clientes
/*-----
&routines clientes

arc addresscreate xxsrfr line 4 1
nu-min-esq
nu-max-esq
nu-min-dir
nu-max-dir
id-log
end
arc addressbuild xxsrfr line
arc addressmatch xxinfo NOPARSE xxsrfr clientes nu-afastamento # NOREJECTS #
# %squeeze%
nu-virtual
[unquote '']
id-log
[unquote '']
[unquote '']
[unquote '']

&return

/*-----
/* Trata erro na execução
/*-----
&routines su-erro

&severity &error &fail
&s fecha [close -all]
&s data [substr [quote [date -vis]] 1 6] [date -year] as [date -time]
&sys echo Erro no programa su-atualiza-clientes.aml em %data% >
$HOME/ALERTA
&sys cat $HOME/ALERTA >> %pathlog%/[date -year][substr [date -tag] 3 4].log

&r $SIURB_AP/fonte/aml/su/su-erro su-atualiza-clientes

&if [show program] = ARCPLOT &then quit
&if [show program] = ARCEDIT &then quit
&if [show program] = ARC &then quit
&return

```

Anexo 6 – Números e indicadores da COPASA ⁷⁵

Abastecimento de Água	Números
Nº de Municípios com Concessões de Serviços *	606
Nº de Municípios com Prestação de Serviços *	569
Localidades com Concessões *	1.018
Localidades Operadas *	820
População Atendida (1.000 habitantes)	11.092
Cobertura em Relação à População Total do Estado (%)	57,2
Unidades Abastecidas (economias faturadas)	3.594.807
Extensão de Redes (Km)	37.275

* Inclui sedes, vilas, povoados e outros - Posição: Novembro/2005

Esgotamento Sanitário	Números
Nº de Municípios com Concessões de Serviços *	168
Nº de Municípios com Prestação de Serviços *	78
Localidades com Concessões *	297
Localidades Operadas *	104
População Atendida (1.000 habitantes)	5.570
Cobertura em Relação à População Total do Estado (%)	28,7
Unidades Abastecidas (economias faturadas)	1.806.840
Extensão de Redes (Km)	11.644

* Inclui sedes, vilas, povoados e outros - Posição: Novembro/2005

⁷⁵ Fonte: COPASA, 2005

Indicadores da empresa

Indicadores Operacionais	Números
Nº de empregados	10.828
Empregado/1.000 ligações (A+E)	2,55
Volume produzido de água (1.000 m ³) *	67.075
Volume consumido de água (1.000 m ³) *	43.973
Volume faturado (1.000 m ³) *	
Água	51.447
Esgoto	26.535
Total	77.982
Perda faturada (%) *	23,31
Perda na distribuição de água (m ³ /econ/mês) *	6,47
Índice de macromedicação (%)	94,69
Índice de hidrometração	99,75
Índice de tratamento de esgoto (%) (1)	27,06

* Média mensal dos últimos 12 meses

(1) Relação entre o volume de esgoto tratado e o volume de esgoto coletado.

Posição: Novembro/2005