

**Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Curso de Pós-graduação em Ciência da Computação**

**Georreferenciamento de Relações Espaciais
Presentes na *WEB*: um Estudo de Caso**

**Disciplina: Bancos de Dados Geográficos
Professor: Gilberto Câmara
Alunos: Guilherme Tavares de Assis
Jean Wanderlei Alves de Oliveira
Colaboradora: Karla Albuquerque de V. Borges**

**Belo Horizonte - MG
Setembro/2003**

1 INTRODUÇÃO

Uma das formas mais naturais de se descrever uma localização é através do uso de referências relacionadas ao local desejado. Por exemplo, se alguém pega um táxi e menciona o endereço para o qual deseja ir, possivelmente o motorista do táxi perguntará por alguma referência sobre o endereço, como "próximo de" algum lugar, "altura de" algum lugar, "em frente a" algum lugar. Tais referências são consideradas relações espaciais, definindo posições relativas no espaço.

Apesar de não serem precisas, as relações espaciais determinam uma relação no espaço que é mais fácil de ser compreendida e imaginada do que, simplesmente, um endereço exato. Sob o ponto de vista da geocodificação, um endereço exato é a forma mais precisa para se determinar um local mas, no dia a dia, o uso de relações espaciais é muito freqüente. Dessa forma, acredita-se que a interpretação correta de relações espaciais leva a uma geocodificação tão importante quanto a de um endereço exato; inclusive, em algumas situações, as duas representações podem ser complementares.

Assim, o objetivo deste trabalho é avaliar a importância do uso de relações espaciais, relativas a determinados pontos de referência, na descrição de diferentes locais mencionados em páginas da *Web*. Para tanto, as seguintes etapas precisam ser cumpridas:

- a) definição dos pontos de referência da cidade de Belo Horizonte a serem tratados;
- b) definição das relações espaciais a serem geocodificadas;
- c) coleta de páginas da *Web* que mencionam os pontos de referência definidos juntamente com alguma das relações espaciais definidas;
- d) implementação de um programa para extração dos pontos de referência e das relações espaciais utilizados nas páginas coletadas;
- e) experimentação prática do programa implementado para se determinar os tratamentos necessários na extração dos dados;
- f) geocodificação dos dados extraídos, envolvendo, a princípio, uma interpretação geográfica para cada relação espacial e, posteriormente, a visualização de um mapa com os dados extraídos;
- g) análise do mapa obtido, avaliando a importância do uso das relações espaciais definidas.

O restante deste trabalho está organizado como se segue. Na seção 2, são discutidos alguns trabalhos existentes sobre relações espaciais. Na seção 3, são descritas a implementação do programa extrator e a experimentação prática realizada, envolvendo os pontos de referência, as relações espaciais e as páginas definidos. Na seção 4, o processo de geocodificação dos dados extraídos e a análise do mapa obtido são descritos. Na seção 5, são apresentadas as conclusões e especificados os trabalhos futuros relativos ao trabalho em questão.

2 RELAÇÕES ESPACIAIS

Relações espaciais estão presentes tanto em linguagens de consulta espacial quanto em aplicações geográficas. Segundo [PaTh97], a representação e o processamento das relações espaciais são cruciais nas aplicações geográficas porque, freqüentemente, no contexto do espaço geográfico, relações entre entidades espaciais são tão importantes quanto as próprias

entidades. Dependendo do domínio de aplicação, algumas relações espaciais se tornam mais importantes que outras.

Em [EgFr91], as relações espaciais foram agrupadas em três categorias:

- a) topológicas: são aquelas que descrevem os conceitos de vizinhança, incidência, sobreposição, mantendo-se invariante ante às transformações como escala e rotação; "disjunto", "adjacente a" e "dentro de" são exemplos de relações topológicas;
- b) métricas: são consideradas em termos de direções e distâncias, sendo que as relações direcionais são aquelas que descrevem a orientação no espaço como, por exemplo, "norte" e "sul", e as relações de distâncias são aquelas que dependem de definições métricas no sentido de se parametrizar o quanto é perto ou longe como, por exemplo, "perto de" e "longe de"; tal parametrização dependerá das circunstâncias e das entidades geográficas relacionadas;
- c) de ordem: são aquelas que expressam a ordem, total ou parcial, dos objetos espaciais, sendo descritas por preposições como "em frente a", "atrás de", "acima de" e "abaixo de".

[PuEg88] considera mais um tipo de relação, a relação *fuzzy*, exemplificado como "próximo a". Considerando que as relações de distância não são precisas e que, em nossa cultura, preposições como "próximo a" e "perto de" não possuem diferenças muito significativas, as relações de distância podem ser consideradas relações *fuzzy*.

As três categorias básicas de relações são importantes para o armazenamento e recuperação das informações, por fornecerem semântica e consistência geométrica às análises realizadas sobre os objetos geográficos armazenados nos SIGs - Sistemas de Informação Geográficos.

Nos últimos anos, progressos foram feitos na formalização de relações espaciais. No entanto, não existe um consenso quanto a um conjunto mínimo de relações. Segundo [RiSU02], a investigação de relações espaciais é uma tarefa desafiante que envolve um grande esforço em diferentes áreas como, por exemplo, geografia e ciências cognitivas, que desenvolvem conceitos para o tratamento de espaço; assim, uma definição formal de relações espaciais se torna necessária para esclarecer os distintos entendimentos sobre relações espaciais entre as pessoas, e para inferir relações corretas entre objetos espaciais.

Segundo [Fran96], geógrafos utilizam-se extensivamente de "raciocínio espacial" em espaços de larga escala, ou seja, espaços que não podem ser vistos ou compreendidos sob um único ponto de vista. O "raciocínio espacial" diferencia diversas relações espaciais, por exemplo relações topológicas ou métricas, e é tipicamente formalizado usando um sistema de coordenadas cartesianas e álgebra vetorial. Este processamento quantitativo da informação é claramente distinto da forma como as pessoas interpretam relações espaciais. Assim, processos de raciocínio qualitativo tornam-se necessários em, por exemplo, sistemas especialistas espaciais e SIG's. O "raciocínio espacial" qualitativo é amplamente utilizado por seres humanos para entenderem e analisarem um ambiente espacial, quando a informação disponível está na forma qualitativa, como ocorre em documentos textuais.

[EgSh98] desenvolveu um modelo formal que captura detalhes métricos para a descrição de relações espaciais em linguagem natural. Detalhes métricos são expressos como refinamentos das categorias identificadas pelo *9-intersection*, um modelo para relações espaciais

topológicas, e fornecem uma medida mais precisa, do que somente a topologia, para identificar se uma configuração geométrica combina ou não com um termo espacial. Os detalhes métricos são classificados em dois grupos: *splitting ratios* como os valores normalizados de comprimentos e áreas de interseções, e *closeness measures* como as distâncias normalizadas entre partes de objetos distintos. O modelo é usado para determinar qual o melhor termo espacial para um relacionamento entre dois objetos geométricos.

3 EXPERIMENTAÇÃO PRÁTICA

Para a realização dos experimentos práticos, foi implementado um programa extrator (vide anexo I) em Perl - *Practical Extraction and Report Language* [WaCS96]. Tal linguagem foi utilizada por se tratar de uma linguagem para processamento de arquivos textos arbitrários com grande facilidade, extraindo cadeias de caracteres desses arquivos e gerando relatórios baseados nos dados extraídos.

Através de uma pesquisa em páginas da *Web*, foram escolhidos os seguintes pontos de referência, localizados na região centro-sul de Belo Horizonte, para serem utilizados na experimentação prática: Minascentro, Praça da Liberdade, Praça da Savassi, Assembléia Legislativa e *Diamond Mall*. Outros pontos de referência, como Igreja de Lourdes, *Shopping Cidade*, Lagoa da Pampulha e UFMG, também foram cogitados, mas apresentaram poucas menções na *Web*.

Dentre as relações espaciais, as mais encontradas em páginas da *Web*, relacionando-se com os pontos de referência escolhidos, foram:

- a) relações topológicas: "no coração de";
- b) relações de ordem: "em frente de", "atrás de", "ao lado de";
- c) relações de distância: "a aproximadamente (metros/km)", "a tantos (metros/km)";
- d) relações *fuzzy*: "próximo de", "nas proximidades de".

Assim, tais relações foram as escolhidas para serem tratadas neste trabalho.

As páginas da *Web* pesquisadas e coletadas para a experimentação prática, que apresentam tanto os pontos de referência quanto as relações espaciais definidos, foram:

- a) <http://www.fiemg.com.br/cidadania/seminario/turismo.html>;
- b) http://br.dir.yahoo.com/Regional/paises/brasil/estados/minas_gerais/cidades/belo_horizonte/transportes_e_turismo/hospedagem/hoteis;
- c) http://www.amagis.com.br/decisao/decisao04_pag15.htm;
- d) <http://www.aonde.com/indicacao/turismo/55063.htm>;
- e) <http://www.ximenes.com.br>;
- f) <http://www.jornalbalcao.com.br>.

Considerando os pontos de referência, as relações espaciais e as páginas *Web* apresentados, o programa extrator foi executado. O conjunto resultado do programa é formado por dados extraídos com a seguinte estrutura: nome do ponto de referência encontrado, tipo da relação espacial utilizada, texto da página que engloba o ponto de referência e a relação espacial, URL. O anexo II exemplifica uma extração realizada, apresentando uma parte de uma página *Web* coletada para a experimentação prática, juntamente com os dados extraídos da mesma. A FIG.1 apresenta o funcionamento algorítmico principal do programa extrator.

- 1º. Inicializar o conjunto de relações a serem consideradas.
- 2º. Inicializar o conjunto de pontos a serem pesquisados.
- 3º. Ler duas linhas da página analisada.
- 4º. Enquanto não chegar o final da página
 - Verificar a existência de alguma relação e algum ponto nestas duas linhas.
 - Em caso positivo, copiar a relação, o ponto e as linhas no arquivo de saída e buscar outras duas linhas.
 - Caso contrário, buscar uma nova linha.

FIG.1 – Funcionamento algorítmico do programa extrator.

Através dos experimentos práticos, foi observado que algumas melhorias poderiam ser realizadas no programa extrator. Um problema encontrado no algoritmo da FIG.1 é que, se em um par de linhas analisado, existir uma relação espacial que se refere a um ponto de referência qualquer e, neste mesmo par de linhas, existir um ponto de referência entre os pontos escolhidos no trabalho, este par de linhas erroneamente estará no resultado. Uma solução para tal problema seria definir uma janela de proximidade, ou seja, calcular a distância entre a relação espacial e o ponto de referência dentro do texto e, assim, caso essa distância estivesse abaixo de um certo limite definido, o par de linhas seria incluído, com uma maior confiabilidade, dentro do conjunto resultado.

Um outro problema a ser destacado é que, para se realizar a extração dos conjuntos de relações espaciais e pontos de referência de forma completa, deveria existir uma taxonomia que contivesse todas as relações possíveis, ou pelo menos uma boa parcela delas, e também todos os pontos de interesse. Em nosso trabalho, foram considerados apenas quatro pontos de referência, mas poder-se-ia utilizar quantos pontos fossem necessários.

4 PROCESSO DE GEOCODIFICAÇÃO

Para geocodificar os dados extraídos (vide anexo II) pelo programa extrator, foi utilizado o *software MapInfo*. Inicialmente, foi preciso definir uma interpretação semântica geográfica para cada relação espacial tratada neste trabalho, uma vez que em linguagem natural, existem várias maneiras de se referenciar a uma mesma relação espacial. Tais definições encontram-se na tabela TAB.1.

Categoria	Relação Espacial	Interpretação Geométrica	Interpretação Semântica Geográfica
Topológica	no coração de/da	Dentro de	A partir da área de referência, foi traçado um círculo ocupando aproximadamente 2 quarteirões.
De ordem	em frente de/a	No entorno de	Como todas as referências ocupam um quarteirão, as relações de ordem foram representadas da mesma maneira, uma vez que, frente, atrás e ao lado passam a significar a mesma coisa. A partir do limite do objeto de referência foi traçado um polígono incorporando o passeio, a largura da rua e parte do quarteirão de frente.
	atrás de		
	ao lado de		
De distância	a aproximadamente (m/Km)		Distância métrica representada por um raio com o valor dado. Como não se sabe a direção certa, o objeto pode ser lançado aleatoriamente dentro do círculo.
	a tantos (m/Km)		
<i>Fuzzy</i>	Próximo de	Perto de	Foi considerado um raio de 500m a partir do centro do objeto de referência. A área formada pelo círculo representa a relação "perto de".
	nas proximidades de		

TAB.1 – Interpretações geográficas e semânticas das relações espaciais.

Após a definição da interpretação de cada relação, foram localizados todos os objetos de referência e feito o lançamento do contorno de cada um deles (Praça da Liberdade, Praça da Assembléia, Minascentro, Shopping Diamond Mall, Praça da Savassi). As áreas correspondentes a cada relação espacial foram feitas de acordo com a tabela TAB.1. A FIG. 2 mostra o mapa gerado.

Para cada objeto extraído, foi armazenado no banco de dados o nome do objeto, o nome da referência, o tipo de relação e a URL da qual o objeto foi obtido. Os objetos foram então localizados utilizando o tipo da relação; por exemplo, "perto de" ficou aleatoriamente dentro do raio de 500m. Posteriormente, usando o endereço exato de cada extração, os objetos foram localizados de forma precisa, de forma a validar se a interpretação dada a cada relação aproximou-se do real. Apenas não foi possível localizar com precisão o objeto representado pela estrela vermelha devido a falta de endereço disponível. Os outros objetos ficaram dentro da área estimada. Apenas as relações de distância apresentaram um valor menor do que o extraído.

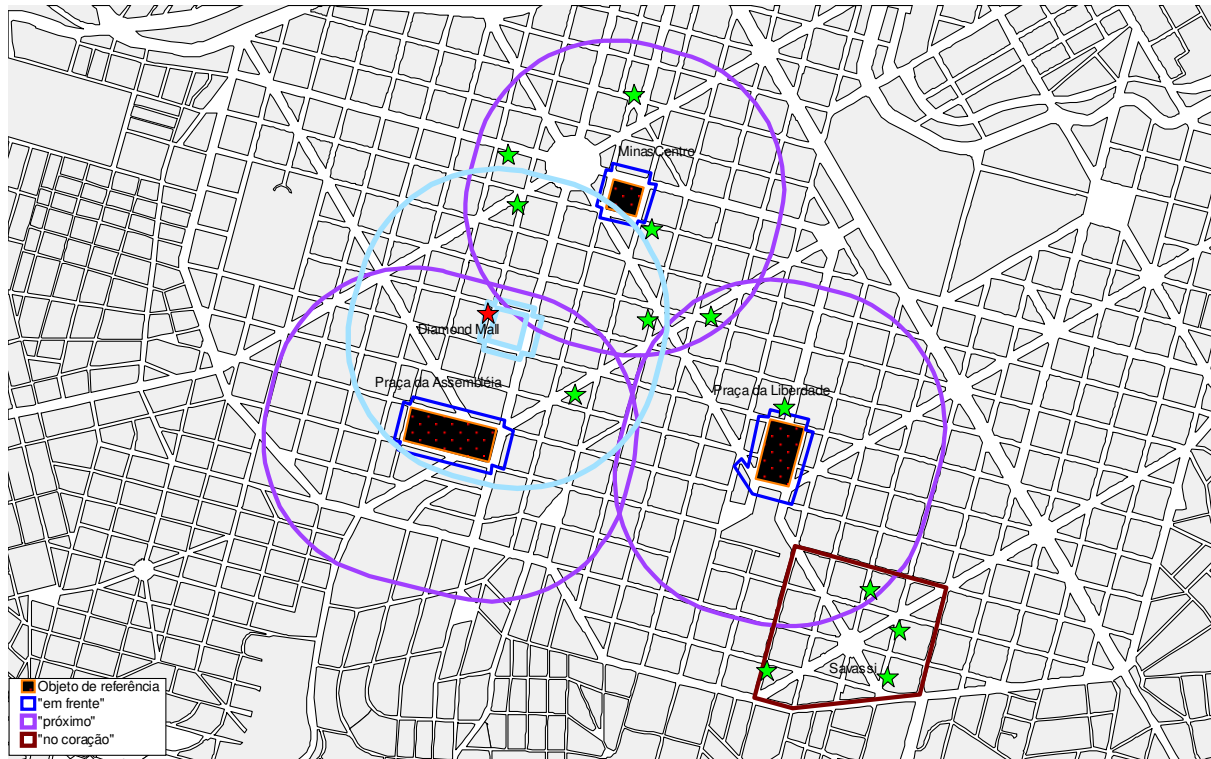


FIG.2 – Delimitação das relações espaciais e localização dos objetos extraídos

A FIG.3 mostra a localização dos objetos extraídos e os pontos de referência relacionados. Por exemplo, todos os objetos representados por um círculo verde fazem referência ao Minascentro.

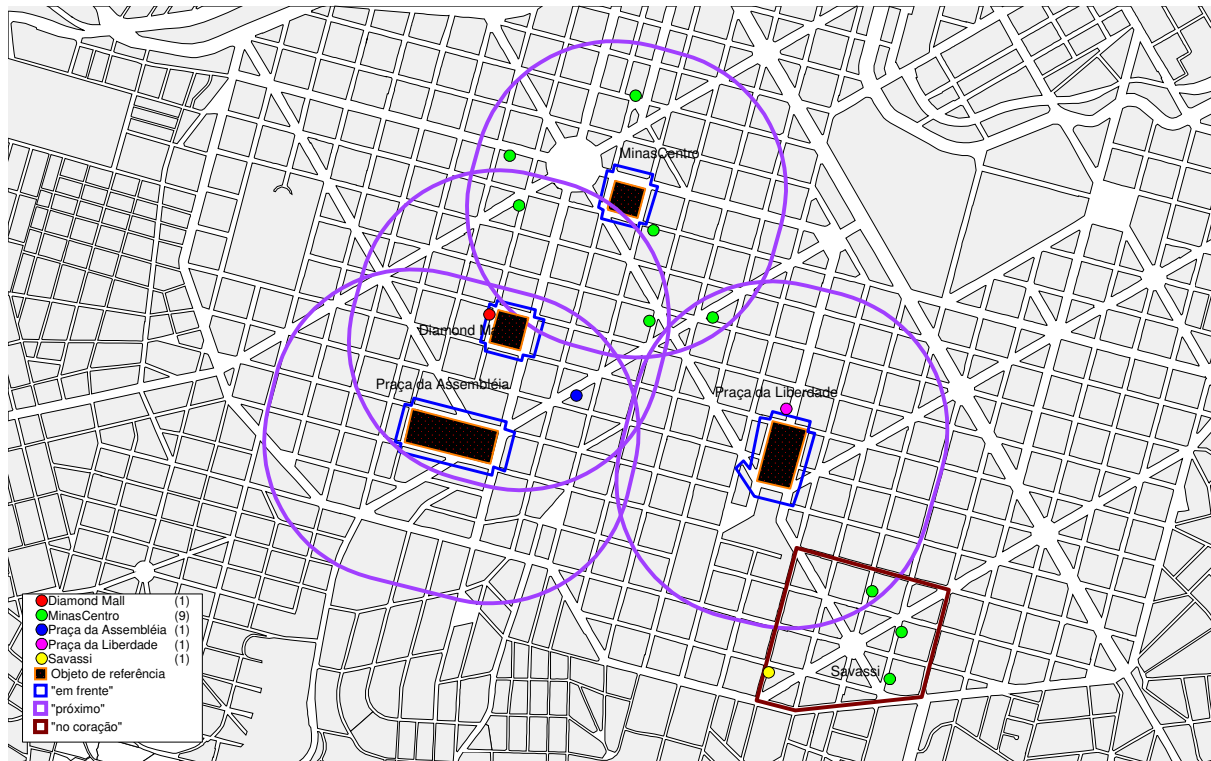


FIG.3 – Objetos segundo o local de referência utilizado.

Já a FIG.4 mostra quais objetos usaram um determinado tipo de relação. Por exemplo, os objetos em amarelo utilizaram a relação "próximo de".

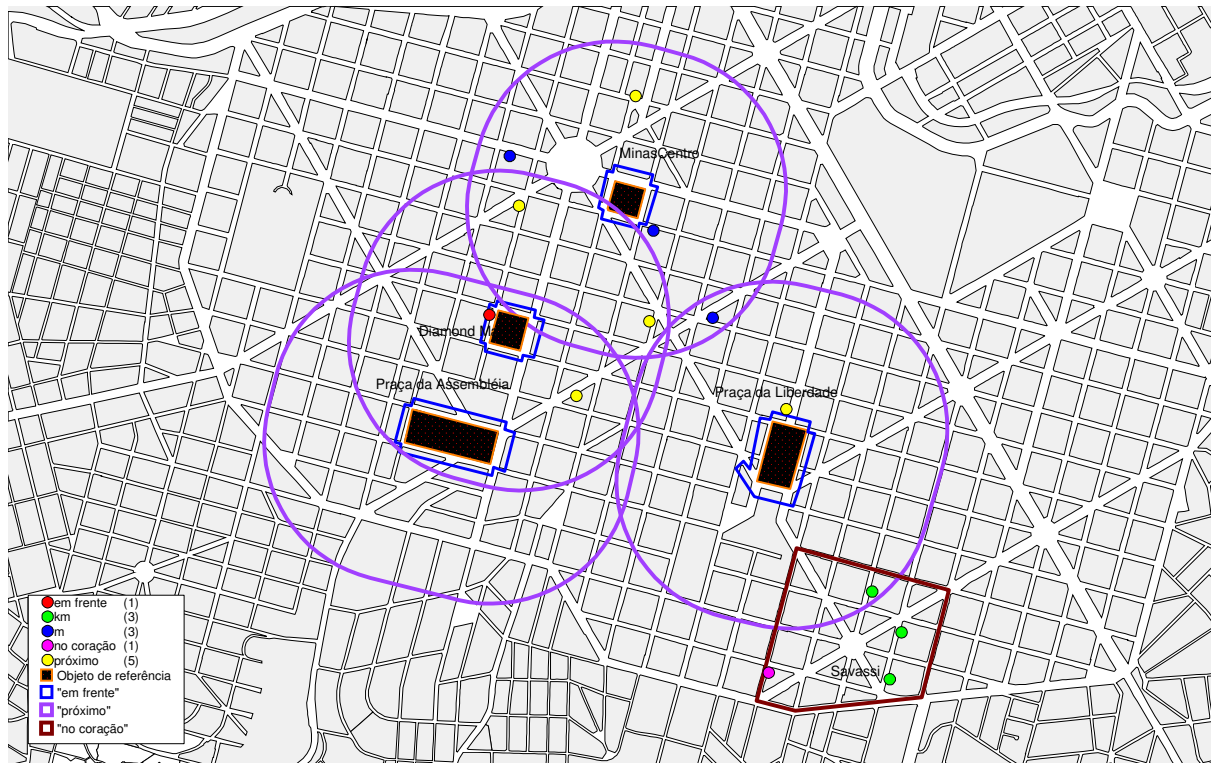


FIG.4 – Objetos extrados segundo o tipo de relao espacial utilizada.

A FIG.5 mostra a combinao do local de referncia e do tipo de relao espacial utilizada por cada objeto extrado. Todos os objetos em verde usaram a referncia "minascentro" e a relao de distncia em km.

Por ltimo, a FIG.6 ilustra a visualizao do mapa no *software MapInfo*, ao se clicar em um determinado objeto georreferenciado. Neste caso, so apresentados os dados referentes ao objeto escolhido, ou seja, o nome do objeto, o nome do ponto de referncia tratado, a relao espacial associada e a URL de onde o objeto foi extrado.

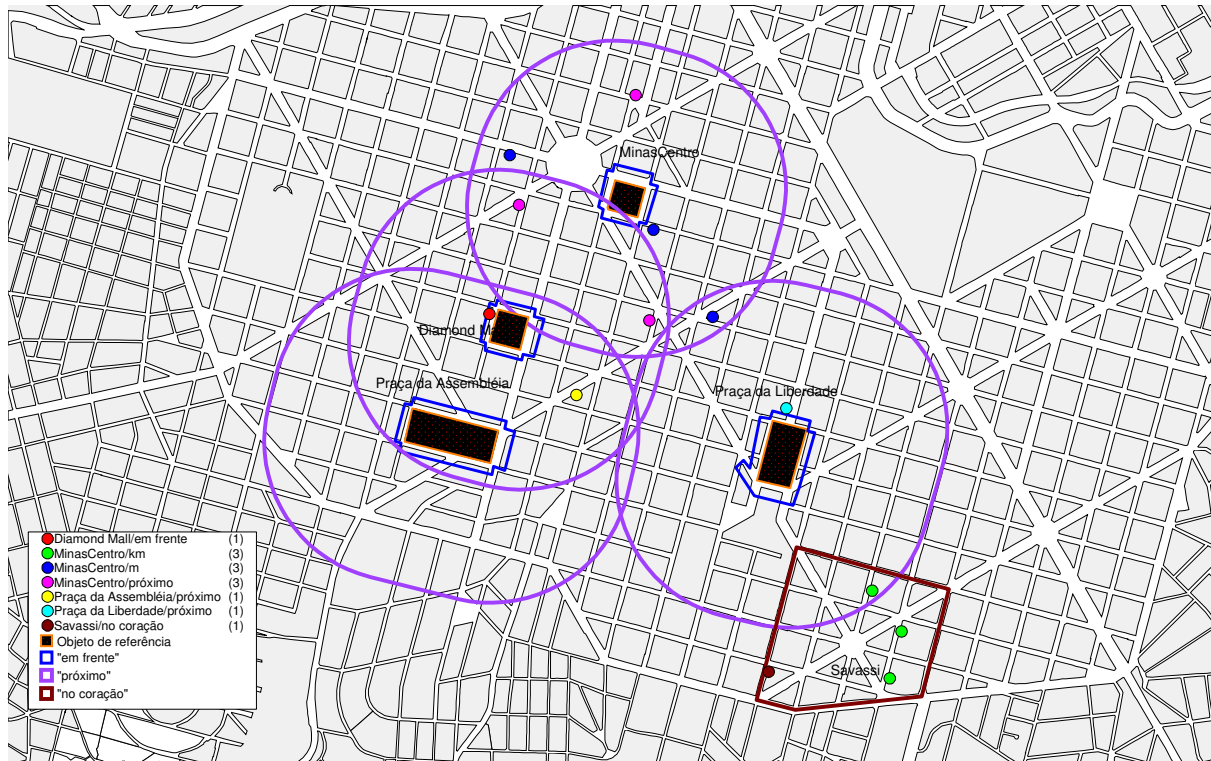


FIG.5 – Objetos extraídos segundo o local de referência e o tipo de relação.

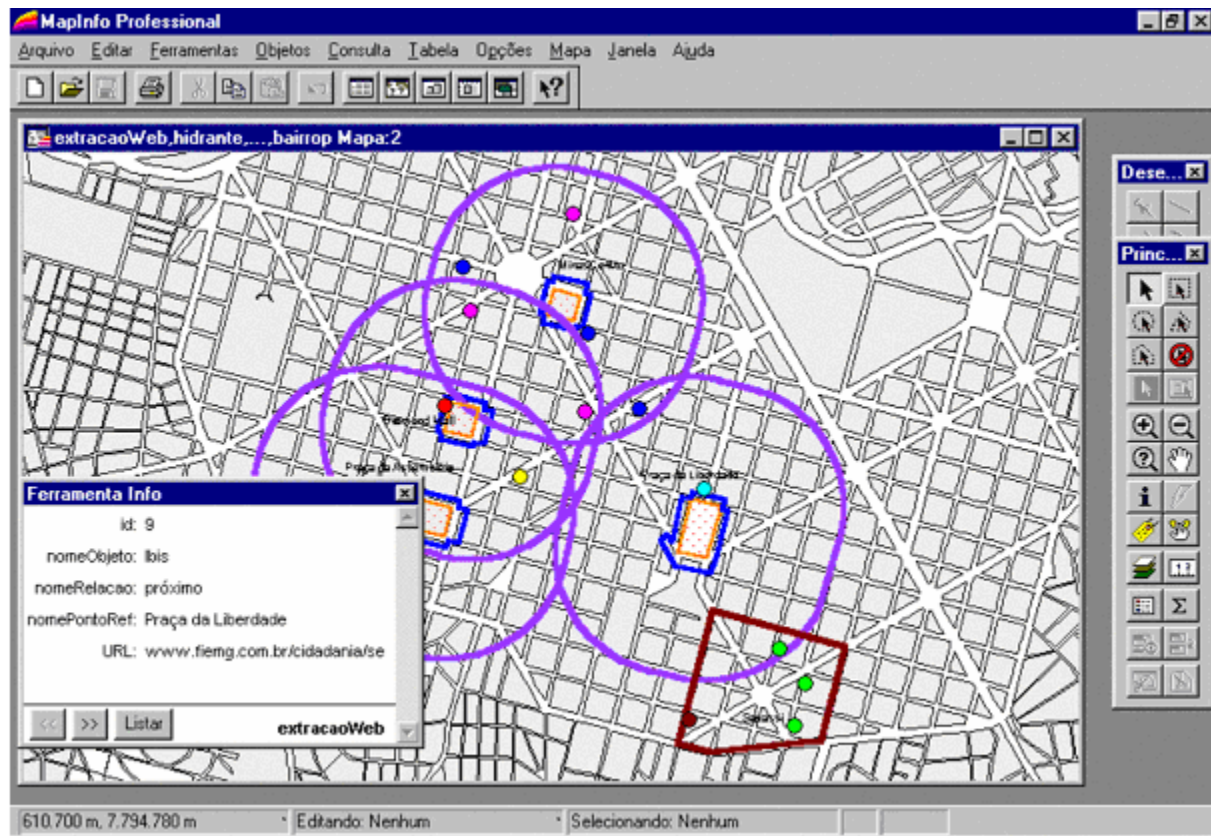


FIG.6 – Escolha de um determinado objeto no MapInfo.

5 CONCLUSÕES E TRABALHOS FUTUROS

Apesar de preliminares, os resultados mostraram que é possível localizar dados extraídos da *Web* utilizando relações espaciais. Existem fontes de dados que não apresentam o endereço na forma completa. No entanto, apesar de um certo grau de imprecisão, os objetos extraídos podem ser localizados dentro de uma determinada região, diminuindo assim a área de incerteza.

O trabalho foi baseado em cinco pontos de referência. Assim, é interessante aplicar o extrator gerado em mais pontos de referência, com o objetivo de aumentar a confiabilidade do mesmo e de avaliar, mais precisamente, a importância do uso de relações espaciais na descrição de localizações.

Além disso, é interessante que o extrator manipule outras relações espaciais, além daquelas tratadas neste trabalho, englobando assim todas as categorias existentes. Dessa forma, poder-se-ia realizar uma análise comparativa mais detalhada entre as categorias de relações espaciais, comprovando qual é a melhor relação espacial e/ou categoria para se descrever uma determinada localização.

Outras geocodificações poderiam ser realizadas, utilizando interpretações semânticas e geográficas para as relações espaciais diferentes daquelas definidas neste trabalho, com o objetivo de se escolher qual é a interpretação mais adequada para uma determinada relação espacial.

Nesse trabalho, foi tratada apenas uma relação espacial e um local de referência por objeto extraído; no entanto, a combinação de mais de uma relação e/ou local de referência (quando existir) é uma forma de refinar a localização. Podemos citar, como exemplo, parte de uma extração: "próximo" da "Praça da Liberdade", sede do governo. "Aproximadamente 1,5 km" do "Minascentro".

Outra aplicação é selecionar, através dos resultados da extração, quais são as referências mais utilizadas na *Web*. Um software de visualização geográfica pode fazer uso desse resultado e predefinir áreas de visualização que contenham esses pontos de referência mais citados.

ANEXO I: CÓDIGO FONTE DO PROGRAMA EXTRATOR

Arquivo "Init.pm"

```
package Init;
require 5.000;
require Exporter;

@ISA = qw(Exporter);
@EXPORT = qw(inithash, initarray);

#!/usr/bin/perl

sub inithash{
    my %hash;
    %hash = (0, "proximo",
            1, "frente",
            2, "lado",
            4, "no coracao",
            5, "nas proximidades",
            6, "a aproximadamente",
            7, "metros",
            8, " m ",
            9, "km",
            10, "atras",
    );
    return %hash;
}

sub initarray{
    my @lista;
    @lista = ('proximo', 'frente', 'lado', 'no coracao', 'nas
proximidades', 'a aproximadamente', 'metros', ' m ', 'km', 'atras');
    return @lista;
}

sub initvetsymbols{
    my @vet;
    @vet = ("À", "Á", "Â", "Ã", "Ä", "Ç", "È", "É", "Ê", "Ë", "Ì", "Í", "Î",
"Ï", "Ñ", "Ò", "Ó", "Ô", "Õ", "Ö", "Ø", "Ù", "Ú", "Û", "Ü", "Ý", "à", "á",
"â", "ã", "ä", "å", "ç", "è", "é", "ê", "ë", "ì", "í", "î", "ï", "ñ", "ò",
"ó", "ô", "õ", "ö", "ø", "ù", "ú", "û", "ü", "ý", "ÿ");
    return @vet;
}

sub initvetcorresp{
    my @vet;
    @vet =
("a", "a", "a", "a", "a", "c", "e", "e", "e", "e", "i", "i", "i", "i", "i", "n", "o", "o", "o",
"o", "o", "o", "u", "u", "u", "u", "y", "a", "a", "a", "a", "a", "a", "a", "c",
"e", "e", "e", "e", "i", "i", "i", "i", "n", "o", "o", "o", "o", "o", "o",
"u", "u", "u", "u", "y", "y");
    return @vet;
}

sub initpontos{
    my @vet;
    @vet = ("praca da liberdade", "assembleia", "diamond mall", "minas
centro", "minascentro");
    return @vet;
}
```

Arquivo "Acha.pl"

```
#!/usr/bin/perl
require "Init.pm";

%refhash = Init::inithash();
@refarray = Init::initarray();
@vetacentos = Init::initvetsymbols();
@vetsemacentos = Init::initvetcorresp();
@pontos = Init::initpontos();

print OUT "$ARGV[0]\n";

if ($ARGV[0] =~ /www(.*)/){
    $url = "http:\\\\\\".$ARGV[0];
    print "$url\n";
    $entrada = `lynx "$ARGV[0]" -source >> entrada.dat`;
    $entrada = "entrada.dat";
}
else{
    $entrada = $ARGV[0];
}

open(IN, $entrada) || die "Invalid file name";
open(OUT, '>saida.dat');

print OUT "$ARGV[0]\n\n\n";

$linha1 = '';
$linha2 = '';

while(chop($linhatemp = <IN>)){ #Lê linha por linha do arquivo de
entrada
    $linha2 = lc($linhatemp);

    for ($i=0; $i<=$#vetacentos; $i++){
        if ($linha2 =~ /(.*?)$vetacentos[$i](.*)/){
            $linha2 = $1.$vetsemacentos[$i].$2;
            $i = 0;
        }
    }
}
```

```
$linhatemp = $linha1.$linha2;

for ($t = 0; $t<= $#pontos; $t++){
    if ($linhatemp =~ /(.*?)$pontos[$t](.*)/){
        $i=0;
        while (($linhatemp !~ /(.*?)$refarray[$i](.*)/ ) &&
($i<$#refarray+1)) {
            $i++;
        }

        if ($i<=$#refarray){
            print OUT "$pontos[$t]\n";
            print OUT "$refarray[$i]\n";
            print OUT "$linhatemp\n\n";
            $linha1 = '';
            $linha2 = '';
        }
    }
}

$linha1 = $linha2;
}
```

ANEXO II: EXEMPLO DE RESULTADO OBTIDO PELO PROGRAMA EXTRATOR

A FIG.7 apresenta uma parte da página “<http://www.fiemg.com.br/cidadania/seminario/turismo.html>” utilizada na experimentação prática.

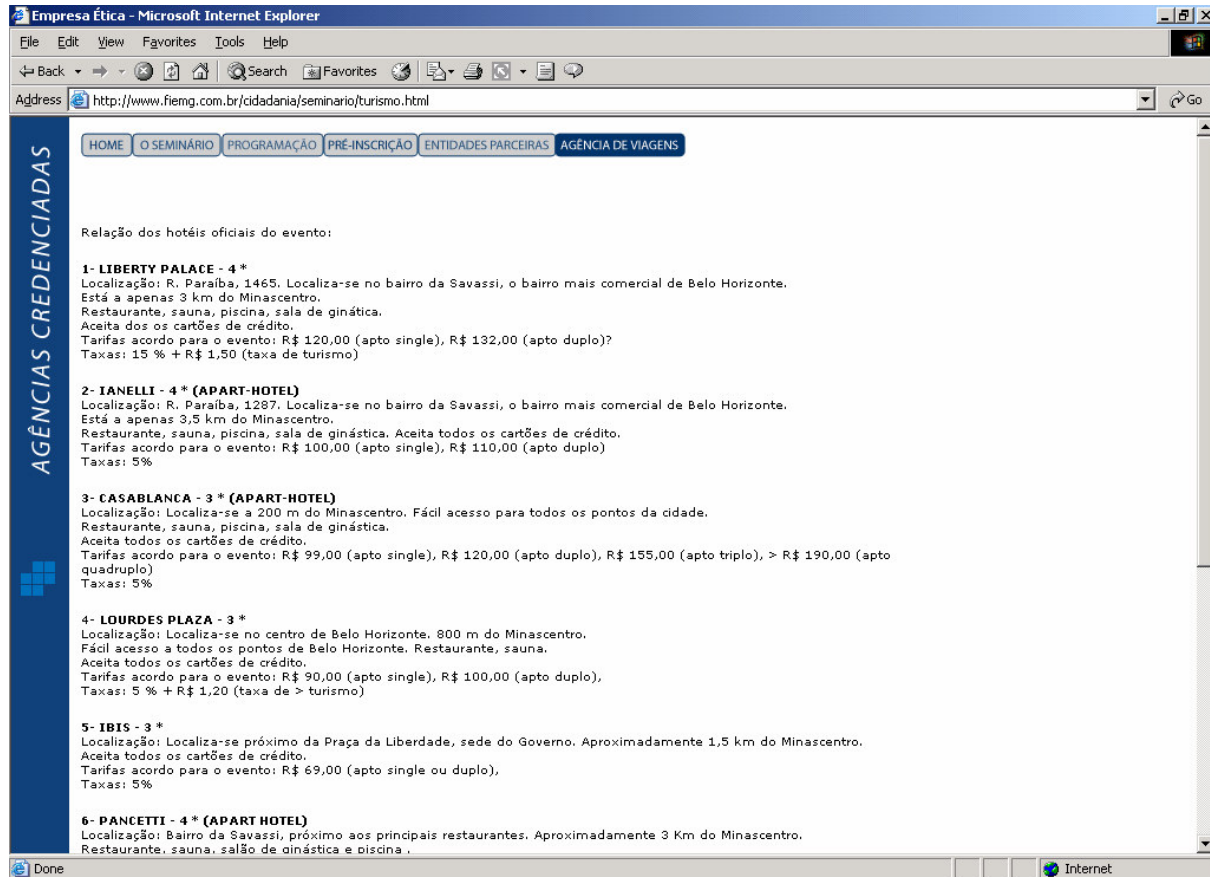


FIG.7 – Página <http://www.fiemg.com.br/cidadania/seminario/turismo.html>

De acordo com a FIG.7, o programa extrator gerou o seguinte resultado:

- Ponto de referência: minascentro
Tipo da relação espacial: km
Texto da página: mais comercial de belo horizonte.
 esta a apenas 3 km do minascentro.

- Ponto de referência: minascentro
Tipo da relação espacial: km
Texto da página: mais comercial de belo horizonte.
 esta a apenas 3,5 km do minascentro.

- Ponto de referência: minascentro
Tipo da relação espacial: m
Texto da página: - 3 * (apart-hotel)
 localizacao: localiza-se a 200 m do minascentro. facil acesso para todos

d) Ponto de referência: minascentro

Tipo da relação espacial: m

Texto da página: plaza - 3 *
 localizacao: localiza-se no centro de belo horizonte. 800 m do minascentro.

e) Ponto de referência: praca da liberdade

Tipo da relação espacial: proximo

Texto da página: - 3 *
 localizacao: localiza-se proximo da praca da liberdade, sede do governo.

BIBLIOGRAFIA

- [EgFr91] EGENHOFER, Max J., FRANZOSA, Robert D. *Point-set topological spatial relations*. International Journal of Geographical Information Systems, London, v.5, n.2, p.161-174, 1991.
- [EgSh98] EGENHOFER, Max J., SHARIFF, A. Rashid B. M. *Metric details for natural-language spatial relations*. ACM Transactions on Information Systems , v.16 , n.4, p.295-321, 1998.
- [Fran96] FRANK, Andrew U. *Qualitative spatial reasonig: cardinal direction as an example*. International Journal of Geographical Information Systems, v.10, n.3, p.269-290, 1996.
- [PaTh97] PAPADIAS, Dimitris, THEODORIDIS, Yannis. *Spatial relations, minimum bounding rectangles, and spatial data structures*. International Journal of Geographical Information Systems, London, v.11, n.2, p.111-138, 1997.
- [PuEg88] PULLAR, D. V., EGENHOFER, M. J. *Towards the defaction and use of topological relations among spatial objects*. In: 3rd International Symposium on Spatial Data Handling (Columbus: International Geographical Union), 1988. Proceedings...p.225-242.
- [RiSV02] RIGAUX, Philippe, SCHOLL, Michel, VOISARD, Agnès. *Spatial Databases with Application to GIS*. San Francisco, CA: Morgan Kaufmann Publishers, 2002. 410p.
- [WaCS96] WALL, L., CHRISTIANSEN, T., SCHWARTZ, R. L. *Programming Perl*, 2nd Edition, CA: O'Reilly & Associates Inc., 1996.