

COMO NÃO SE PERDER COM 9 MILHÕES DE ENDEREÇOS

Francisco Aranha

Professor do Departamento de Informática e Métodos
Quantitativos da EAESP/FGV e Consultor de Empresas.

E-mail: faranha@eaesp.fgvsp.br

Resumo: Na adoção de novas tecnologias, o problema nem sempre é a disponibilidade de recursos, mas sim a possibilidade de acesso à informação adequada e à capacidade de utilizar novos quadros de referência analíticos. Quando se empregam Sistemas Geográficos de Informação no tratamento de endereços para o *Database Marketing*, é preciso passar a enxergar os endereços como as entidades autônomas que realmente são, em vez de considerá-los como meros atributos dos clientes. Essa estratégia reduz, consideravelmente, o esforço de geodecodificação.

Abstract: *When adopting new technologies, the issue is not always the availability of resources, but the possibility of access to adequate information and the ability to use new analytical reference sets. When Geographic Information Systems are applied to the treatment of addresses for Database Marketing, it is necessary to view addresses as the autonomous entities they really are, rather than consider them mere client attributes. This line of action greatly reduces the effort involved in address matching.*

PALAVRAS-CHAVE: Sistemas Geográficos de Informação (GIS), *marketing* de relacionamento, geocodificação de endereços, *marketing* direto, listas de endereçamento, banco de dados relacionais.

KEY WORDS: *Geographic Information Systems (GIS), database marketing, geocoding address, direct marketing, mailing lists, relational databases.*

DOIS DESCONHECIDOS COM MUITO EM COMUM

Abordamos recentemente, em outro artigo, as afinidades existentes entre *Database Marketing* e Sistemas Geográficos de Informação:¹ se, de um lado, a atividade de *Marketing* de Relacionamento consiste em um processo de comunicação documentado e de mão dupla entre empresa e mercado consumidor, de outro, os Sistemas de Informação Geográfica oferecem ferramentas que permitem a inclusão de variáveis de localização na análise dos dados recolhidos na interação com o mercado.

Ambos, atividade e ferramenta, estão baseados na manipulação de grandes volumes de informação qualificada e detalhada. Ambos, por conta da novidade e sofisticação da tecnologia envolvida, são bastante desconhecidos; e não apenas do público em geral, mas também dos profissionais que atuam em cada uma das duas áreas.

Em pesquisa que realizamos com 120 encarregados, gerentes e diretores de *Marketing*, entre os meses de junho e agosto deste ano, constatamos que apenas 9 dos entrevistados tinham ouvido falar de GIS. Por outro lado, os mercadólogos também reclamam de ter que enfrentar o que Wilmar Munhós, presidente do IBD, Instituto Brasileiro de *Database Marketing*, delicadamente chama de "cultura de desconhecimento de *Database Marketing*".²

Essa situação deve mudar rapidamente, já que, em ambos os campos, temos tido movimentações importantes. Por exemplo, com a finalidade de explorar um banco de dados com mais de 9 milhões de nomes,³ segmentados em 22 categorias de consumidores, o Grupo Abril lançou este ano a DataListas, empresa formada em *joint venture* com a US West Inc, uma das primeiras do mundo em telecomunicações, entretenimento em TV a cabo e sistemas de informação. De outro lado, a ESRI, líder mundial em tecnologias de geoprocessamento, comprou a Strategic Mapping Inc., fabricante do Atlas GIS e líder no segmento de *desktop mapping*. O reflexo no Brasil foi uma dinamização dos dois mercados.

Mesmo a mídia não especializada começa a perceber toda essa agitação. O jornal *O Estado de São Paulo* publicou, em 9 de se-

tembro, matéria sobre o lançamento, em português, do MapInfo, um dos principais softwares de *desktop mapping*⁴ e a revista *Exame* da primeira quinzena do mês publicou uma matéria sobre a aplicação de GIS à análise de pontos comerciais.⁵ Mais exposição desses temas nos meios de comunicação vai provocar interesses e estimular trocas de informação não apenas entre as duas áreas, mas também entre usuários potenciais, prestadores de serviços e fornecedores de *soft* e *hardware*. É um mercado em aquecimento.

DATABASE MARKETING E GIS ESTÃO BASEADOS NA MANIPULAÇÃO DE GRANDES VOLUMES DE INFORMAÇÃO QUALIFICADA E DETALHADA. AMBOS, POR CONTA DA NOVIDADE E SOFISTICAÇÃO DA TECNOLOGIA ENVOLVIDA, SÃO BASTANTE DESCONHECIDOS.

DE APALPADELAS E PERPLEXIDADES

Esse intercâmbio de informações e experiências possibilita um reconhecimento do terreno e acaba levantando questões interessantes.

Em uma reunião de trabalho da qual participamos no final de agosto e em que se discutia o potencial de aplicação de GIS ao *Database Marketing*, defendíamos a tese de que os administradores de listas deveriam incluir, urgentemente, em seus bancos de dados, dois campos numéricos e ali armazenar a latitude e a longitude de cada endereço.

Com essa simples providência, os endereços passariam a ser prontamente localizáveis em mapas, permitindo uma enorme variedade de formas de pesquisa e seleção baseadas na posição geográfica da informação. Os indivíduos constantes do banco de dados poderiam ser escolhidos não apenas por suas características sócio-econômicas, mas também por localização, até mesmo de quarteirão, se fosse o caso. Questões do tipo: "gostaria de saber o nome e en-

1. ARANHA, Francisco. Sistemas de informação geográfica: uma arma estratégica para o *database marketing*. In: *RAE - Revista de Administração de Empresas*, EAESP-FGV, São Paulo, V. 36, n.2, p.12-16, Abr./Maio/Jun. 1996.

2. QUEREMOS mais que um mailing. In: *Netpeople*, Datalistas, Ano I, 1996, p.3.

3. DATALISTAS inicia operações com 9 milhões de nomes no banco de dados. In: *Meio e Mensagem*, 12 ago. 1996.

4. MAPINFO em português traz novos recursos. In: *O Estado de São Paulo*, 02 set. 1996.

5. EXAME. Esses programas dão o mapa da mina. São Paulo: Abril, Edição 618, ano 30, n.19, p.90-1, 11 set. 1996.

dereço dos chefes de famílias de classe B e C com crianças, residentes em um raio de 5 Km da nossa escola" não apresentariam nenhuma dificuldade. A pesquisa poderia ser feita diretamente no banco de dados, por meio de um simples comando de busca.

A geocodificação, isto é, o processo que associa as coordenadas geográficas de um endereço ao seu registro na *database*, não oferece complicações teóricas. A latitude e a longitude de um imóvel podem ser estimadas aproximadamente a partir de pontos de referência com localização conhecida, como, por exemplo, o ponto médio de um segmento de rua (essa informação pode ser extraída com facilidade de um mapa digital de ruas); ou o centróide de um setor postal; ou diretamente de sinais de satélite do sistema GPS.⁶

Quando estamos interessados em localizar, por exemplo, os postos de gasolina ao longo de uma avenida, ou os restaurantes de um bairro, basta percorrer a região com um rastreador de satélite em mãos (hoje podem ser do tamanho de uma calculadora de bolso), apertar uma tecla cada vez que passarmos em frente de um posto ou um restaurante e depois descarregar os registros de coordenadas no computador. Um GIS coloca os pontos diretamente no mapa.

Nossa tese a respeito da conveniência da geocodificação foi recebida com resistência. A objeção levantada naquela ocasião foi a seguinte: geocodificar **alguns poucos** endereços realmente é fácil, mesmo que tenhamos que buscar coordenadas de um satélite. Geocodificar **nove milhões** de endereços é um problema bastante distinto. As dificuldades começam com o tamanho brutal da tarefa. Nove milhões de endereços correspondem a mais que o dobro de **todos** os domicílios dos 39 municípios da Grande São Paulo. A geocodificação deveria, necessariamente, ser realizada por um processo automatizado.

Até recentemente, não tínhamos encontrado uma resposta a essa objeção. Mas o esforço de visualizar a quantidade "nove milhões provocou-nos um estalo.

ENDEREÇO É ENTIDADE, NÃO ATRIBUTO

Nove milhões de nomes não correspondem a nove milhões de endereços.⁷ E mesmo nove milhões de endereços não correspondem a nove milhões de lugares diferentes, principalmente se pensarmos em termos de grandes centros urbanos. Em uma cidade como São Paulo, por exemplo, cada edifício representa cerca de, digamos, 60 apartamentos, todos literalmente "empilhados" na mesma localização. Geocodificar um apartamento equivale a geocodificar todos os demais no mesmo edifício. O mesmo acontece com as lojas de um *shopping center*, ou os escritórios de um centro comercial.

Essa constatação traz o problema para dentro do campo da análise geográfica. E não estamos muito habituados a esse tipo de raciocínio; ao contrário do que nos pareceu inicialmente, para geocodificar os nomes de um banco de dados daquela dimensão, não procuraremos uma quantidade correspondente de lugares distintos. Em vez disso, dividiremos o procedimento em duas etapas: primeiro geocodificar os lugares e depois relacionar as pessoas com eles.

Para isso é necessário desfazer uma confusão consolidada: endereço não é atributo de pessoas, mas entidade autônoma, isto é, o lugar onde um consumidor mora; não é uma característica própria dele, consumidor, como a idade ou o número de filhos. Um indivíduo pode mudar de casa, e a casa continua lá, no mesmo endereço, com o mesmo CEP, as mesmas coordenadas geográficas; portanto ocupando o mesmo lugar no mapa.

Em conseqüência, em um banco de dados relacional, os nomes deveriam formar uma tabela, e os endereços, outra.

Na tabela de segmentos de ruas (ou faces de quadra), os registros conteriam, por exemplo:

- o tipo (rua, avenida, travessa etc.) e nome atual do logradouro;

6. *Global Positioning System: uso de sinais de satélite para levantamentos, medições de precisão, navegação e determinação de altitudes.* Um bom livro sobre o tema é HOFMANN-WELENHOF et alli. *Global positioning system - theory and practice*, 3 ed., Springer-Verlog, 1994, 326 p.

7. Disso decorre, por exemplo, a necessidade de um serviço chamado *household analysis*, em que se procura selecionar apenas um destinatário de um determinado domicílio, para evitar desperdício de recursos em campanhas de mala direta.

Cód-log	Quadra	Tipo	Título	Prep.	Nome	Inicial	Final	Inicial	Final	Direita	Esquerda	Lon	Lat
171.212-0	1	R	Br	do	Rio Branco	14	52	43	69	04753-000	04753-000	32.5822	73.8294

- nomes anteriores (já que, infelizmente, neste país os legisladores municipais dedicam uma parte significativa de seus projetos a mudar nomes de ruas, praças e viadutos);
- numeração inicial e final, à direita e à esquerda;
- o CEP ou CEPs que lhe correspondem;
- suas coordenadas geográficas.

Na tabela de clientes, teríamos seus dados pessoais e, no campo de endereço, apenas:

Logradouro	Quadra	Número	Complemento
171,121-0	1	22	conj 13

Esse procedimento tem enormes vantagens:

- evita a repetição do nome da rua nos registros de clientes, gerando economia de espaço no disco rígido;
- permite que o endereço seja consistido quando o cliente é cadastrado, pois o nome da rua, número do imóvel e seu CEP devem ser coerentes com o correspondente registro de segmento de rua na tabela de rua;
- padroniza a forma de anotação dos endereços, evitando confusões (quem já trabalhou com malas diretas sabe o problema que é encontrar em um banco de dados não sistematizado os clientes cujos escritórios estão, por exemplo, na Rua Barão do Rio Branco, simplesmente porque eles podem ter sido registrados na:

R	Br	do	Rio Branco
Rua	Barão do		Rio Branco
R.	Brão do		Rio Branco
R.	Br.	de	R. Branco
Rua			Rio Branco
R.	Barão do		R. Bco
		etc)	

- facilita atualizações como, por exemplo, quando a rua muda de nome, e
- permite, pela utilização das coordenadas do ponto médio do segmento, que o endereço seja prontamente localizável em um mapa, com aproximação bastante razoável.

Segundo essa estratégia, a tarefa aparentemente gigantesca de geocodificar, por exemplo, todos os cerca de 10 milhões de habitantes do município de São Paulo fica reduzida ao cadastramento dos seus "apenas" 145 mil segmentos de rua. No caso particular da capital paulista, um arquivo de ruas bastante razoável pode ser adquirido de diversas fontes, por um custo relativamente modesto, e ocupa cerca de 35 megabytes de disco rígido. Em outras palavras, cabe em qualquer computador e no bolso de qualquer empresa.

DISPOSIÇÃO PARA MUDAR DE PERSPECTIVA

Em resumo, quando exploramos novas tecnologias, como nesse caso da aplicação de Sistemas Geográficos de Informação ao *Database Marketing*, uma parte significativa das dificuldades na incorporação dos avanços tecnológicos vem da própria novidade do tema. Às vezes, o recurso já está disponível, mas falta informação sobre ele, ou falta mudar o quadro de referência a partir do qual interpretamos a informação disponível.

Para geocodificar nove milhões de endereços, hoje em dia, basta querer e não se perder em velhas estruturas de pensamento. □

Uma trilha nova no mapa da mina.

No número anterior da RAE, demos algumas indicações de leitura e sugestões para auxiliar os interessados na escolha de um programa de *desktop mapping*. No entanto, na briga pela preferência do consumidor, houve um desdobramento importante do qual, até então, não tínhamos conhecimento: foi o lançamento do Maptitude, pela empresa Caliper. Esse aplicativo, que recentemente tivemos a oportunidade de testar, é muito amigável, apresentando uma das interfaces mais intuitivas dentre as que estão disponíveis no mercado; é bastante econômico em termos de espaço em disco; e, por último, mas não menos importante, custa cerca de 30% menos que o próximo concorrente mais barato. Sendo assim, junte essa "dica" às anteriores e considere o Maptitude como uma alternativa, quando estiver se decidindo pela compra do seu *software*.