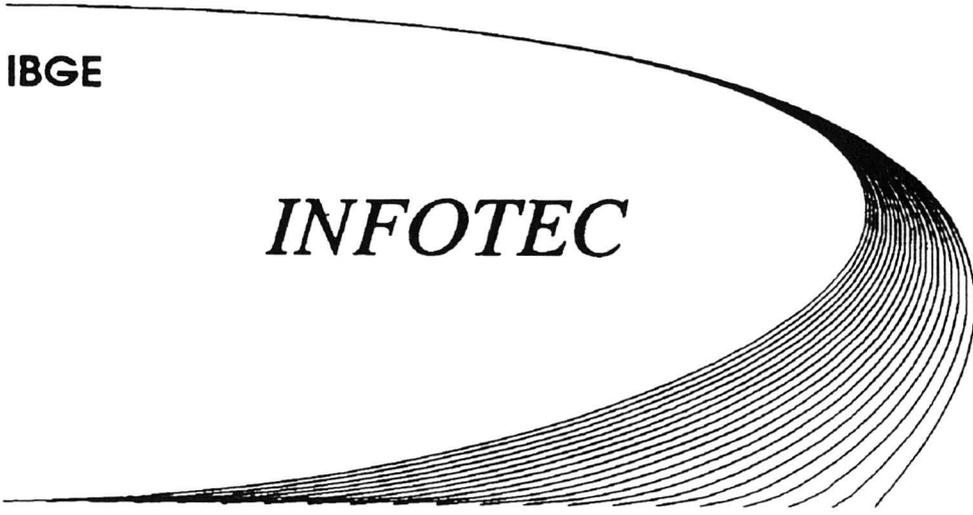


IBGE

INFOTEC



Informativo Técnico DI - Abril 92

Ano I - nº 4

Leia neste número:

CRIPTA

**Uma Ferramenta para Geração Automática
de Sistemas de Crítica e Correção de Dados**

Reina Marta Hanono
DI / DETEC / DISIA

CRIPTA

Uma Ferramenta para Geração Automática de Sistemas de Crítica e Correção de Dados

Resumo

Neste trabalho descrevemos o sistema CRIPTA, Gerador Automático de Programas de Crítica e Correção, sendo desenvolvido no IBGE. Este sistema se propõe a ser uma ferramenta destinada aos usuários não necessariamente especialistas em informática, e se integra no Esquema Genérico para Processamento de Dados Estatísticos, proposto pelo DETEC 'Departamento Técnico da Diretoria de Informática', seguindo as suas diretrizes no que diz respeito à descentralização do processamento das aplicações de crítica e correção de dados, e à geração de aplicações portáteis (CRIPTA V.2).

No momento estão disponíveis:

- CRIPTA V.1 - gera aplicações para ambiente centralizado IBM/MVS, que utilizem o *PL/I Optimizing Compiler*;
- CRIPTA V.2 - gera aplicações de crítica no Sistema Central IBM/MVS em linguagem 'C', para serem executadas no ambiente descentralizado (UNIX,DOS), e aplicações de atualização no sistema descentralizado (UNIX);
- CRIPTA V.3 - proposto para ser implementada em 1992, objetivando a geração das aplicações de crítica no ambiente descentralizado (UNIX).

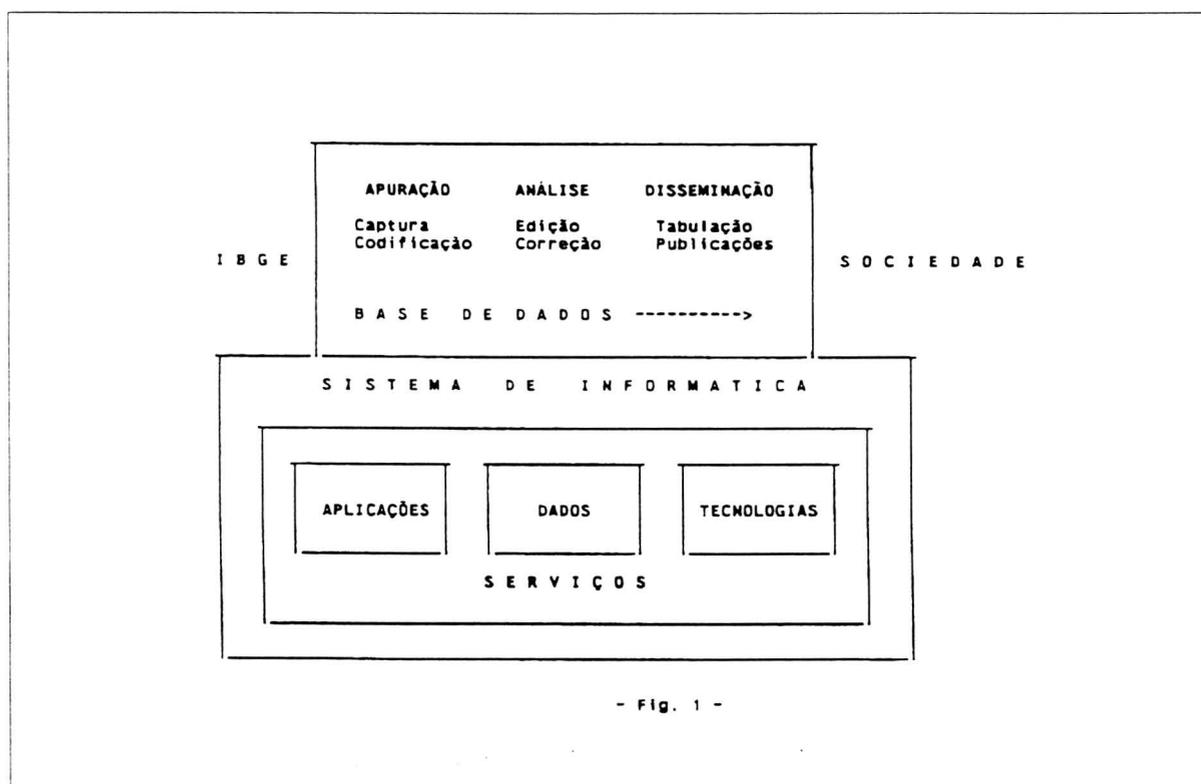
Publicações Anteriores:	disponíveis na Biblioteca DI
INFOTEC 1	CRIPTA Murilo Cardoso de Castro Reina Marta Hanono
INFOTEC 2	O Acervo de Dados do IBGE Mauro Sérgio Cabral
INFOTEC 3	A Rede BitNet Carlos Cardoso da Silva Marcos A. de Oliveira Paulo Paulista Gonçalves

Introdução

Durante a última década a comunidade de informática se orientou à produção de Sistemas de Informática de alta qualidade, baixo custo e bom desempenho.

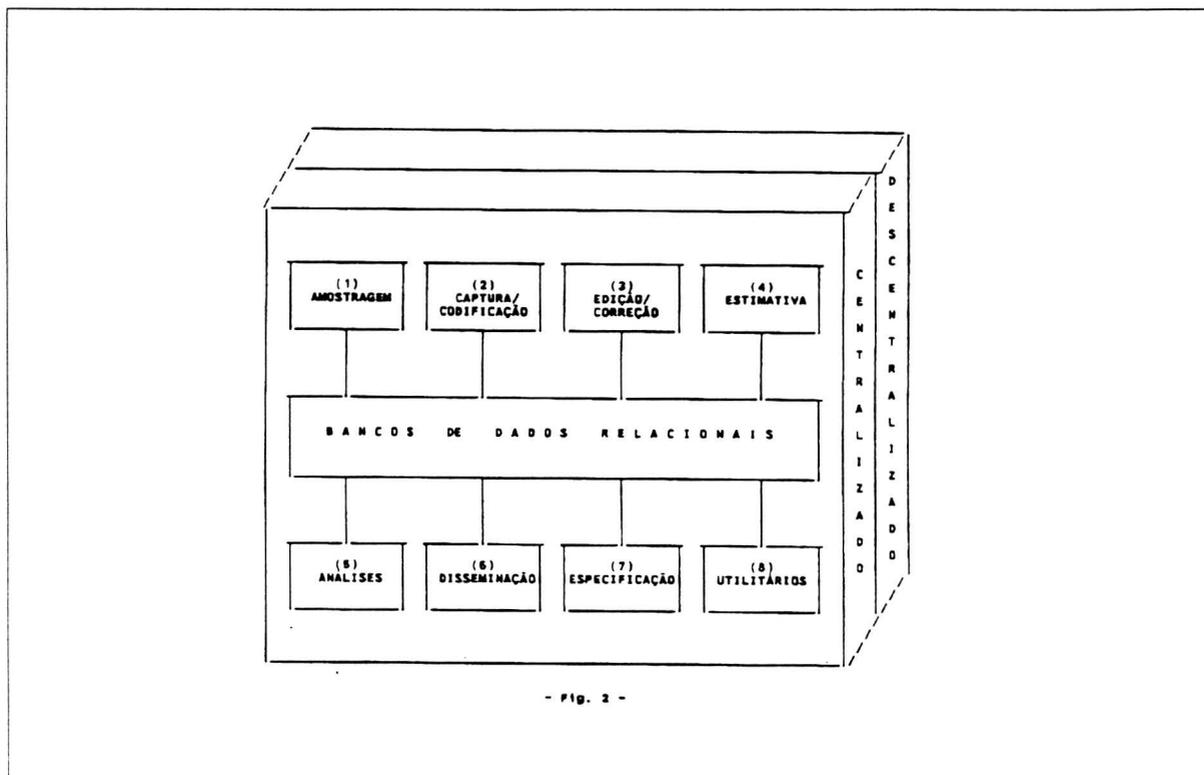
Conscientes destes objetivos, os técnicos de informática do IBGE têm procurado um conjunto de ferramentas para o usuário, de fácil utilização, orientadas a aumentar a qualidade e a produtividade no desenvolvimento e produção de Sistemas de Informática para Processamento de Dados Estatísticos.

Entendemos por Sistemas de Informática para Processamento de Dados Estatísticos um conjunto de aplicações, dados, tecnologias e serviços, que juntos sustentam as diferentes etapas do Ciclo de Vida de uma Pesquisa: Apuração (Coleta, Captura, Codificação e Crítica/Correção), Análise (Crítica, Imputação e Análise de Dados), Disseminação e Acervo (Fig. 1).



Neste sentido o DETEC, Departamento Técnico da Diretoria de Informática, encarregado da questão tecnológica, tem definido uma arquitetura de informática para os anos 90, cujas principais diretrizes são a descentralização da produção das Pesquisas e Censos, a portabilidade das aplicações e o uso de ferramentas genéricas, minimizando o esforço para o desenvolvimento de sistemas de aplicação, propondo a implantação de um **Esquema Genérico para Processamento de Dados Estatísticos**, baseado no **GSFD** (*General Survey Function Design*) do Statistics Canada (Fig. 2).

Visualizando a fase de Crítica/Correção da etapa de Apuração, fizemos um estudo das ferramentas de mercado, assim como de outras instituições de estatística (3) e, sendo que nenhuma delas se adaptava às necessidades por nós impostas, demos partida ao desenvolvimento de uma nova ferramenta chamada 'CRIPTA', que nada mais é que um gerador automático de programas de crítica e correção de dados.



O CRIPTA permite:

- ✓ Gerar o programa da aplicação a partir do Plano de Crítica definido pelo usuário final;
- ✓ Gerar e manter uma Base de Dados, contendo as descrições do arquivo lógico a ser processado e o plano de crítica (no IBGE esta função está totalmente integrada à Base de Metadados da Instituição);
- ✓ Acessar dados residentes em arquivos sequenciais ou em Sistemas Gerenciadores de Banco de Dados (DB2, no sistema central IBM e, TSGBD (19), nos sistemas distribuídos UNIX), sem nenhum esforço para o usuário;
- ✓ Integrar-se com outros sistemas orientados para o processamento de dados estatísticos, tais como: DIA (16), SAS (17), ATLAS (18), IMPS (20) e REDATAM (21);
- ✓ Atender a diferentes plataformas, desde o *mainframe* IBM, até sistemas distribuídos UNIX.

Desta forma, o CRIPTA se torna uma poderosa ferramenta automatizada, que promove uma maior integração entre as equipes desenvolvedoras, propiciando uma efetiva participação dos especialistas temáticos, eliminando os problemas de comunicação e mantendo a integridade da documentação.

Visão Geral do Sistema

O CRIPTA que está sendo desenvolvido no momento, no IBGE, é um conjunto de programas computacionais interligados num sistema, que pode ser usado numa ampla gama de situações, isto é, o sistema não foi feito sob medida para uma aplicação específica.

Embora o desenvolvimento e implementação de um sistema deste tipo seja uma tarefa complexa, a complexidade tem sido reduzida desenvolvendo-se o sistema em módulos, que realizam as distintas tarefas da fase de Crítica e Correção.

Esta modularidade permite um progressivo desenvolvimento do sistema, no sentido de que é possível ampliar as funções do mesmo, incorporando-se novos módulos.

Vale ressaltar que esta modularidade propicia, também, uma fácil manutenção, permitindo a incorporação de melhorias a partir das solicitações dos usuários.

Ambiente Computacional

No momento a **Versão 1** está liberada, e gera aplicações para o sistema central IBM, com o PL/I *Optimizing Compiler*, que podem ser utilizadas em ambientes *batch* ou *on-line*. No caso das aplicações para o ambiente *on-line*, o usuário será responsável pelo *front-end* da aplicação.

A **Versão 2**, que viabiliza a geração automática de aplicações de crítica e correção em linguagem 'C', para estações de trabalho (UNIX, DOS) está hoje sendo testada pela DISIA com a simulação da PNAD. Vale ressaltar que as aplicações são ainda geradas no sistema central IBM.

Ambas versões (1 e 2) interagem com Sistemas Gerenciadores de Banco de Dados, tanto para recuperação dos metadados (IDMS, no sistema central IBM), como para o acesso aos Dados (DB2, no sistema central IBM e, TSGBD, nos sistemas distribuídos UNIX).

Vantagens

Podemos citar as seguintes vantagens do Sistema CRIPTA:

- ✓ permite concentrar o trabalho, tanto quanto possível, no departamento do especialista temático;
- ✓ permite especificar a estrutura do questionário uma única vez;
- ✓ tem interface amigável, tanto para especificação da estrutura do questionário, como para especificação do plano de crítica;
- ✓ permite a geração automática de programas de crítica/correção;
- ✓ possibilita reduzir, tanto quanto possível, a natural reciclagem das etapas de crítica e correção de dados;
- ✓ possibilita a geração automática de programas de imputação tipo *hot-deck* (versão 1);
- ✓ produz estatísticas dos erros;
- ✓ provê a geração automática da função de validação por item para aplicações de correção *on-line* (versão 2);
- ✓ possibilita que a correção possa ser executada num processo interativo entre o especialista temático e o computador;
- ✓ permite checagem automática da estrutura do questionário, assim como a consistência automática das variáveis (versão 1);
- ✓ está preparado para ser usado em pesquisas sociais e econômicas;
- ✓ permite o processamento de variáveis inter-registro;
- ✓ provê um bom conjunto de funções automatizadas para serem usadas durante a crítica, tais como funções lógicas, aritméticas e booleanas;

- ✓ permite interfacear com Sistemas Gerenciadores de Banco de Dados;
- ✓ permite interfacear com distintas ferramentas para análise de dados;
- ✓ possui uma completa documentação de utilização.

Especificação de uma Aplicação em CRIPTA

Para se especificar uma aplicação em CRIPTA, o usuário deve especificar:

➔ O Dicionário da Aplicação:

Este contém o *lay-out* do arquivo lógico a ser processado pelo sistema, assim como a descrição das variáveis de trabalho que serão usadas pela aplicação. Hoje, no IBGE, a forma de se especificar o Dicionário está totalmente integrada à Base de Metadados.

Também é possível especificar-se o Dicionário num arquivo sequencial utilizando-se formulários padrão.

➔ Plano de Crítica:

O Plano de Crítica é especificado através de formulário-padrão, no PCAUT, 'Plano de Crítica Automatizado'. Como o Dicionário da Aplicação, está totalmente integrado à Base de Metadados da Instituição, mas pode, também, ser especificado em arquivo sequencial.

O formulário do PCAUT está definido por campos específicos, nos quais o usuário deve especificar as regras.

As regras de crítica são visualizadas pelo PCAUT como uma concatenação de:

- uma **Referência** que identifica a regra;
- um **Procedimento** onde o usuário deve especificar, em CRIPTA, o procedimento a ser efetuado e a condição a ser checada para verificar a regra;
- um **Efeito** a ação que deve ser tomada pelo sistema em resposta à satisfação ou não da condição especificada no procedimento.

➔ Um Programa CRIPTA:

Este tem estilo pseudo-código, formato livre, e deverá ser composto por blocos de procedimentos, delimitados pelos comandos:

```
PROC <identificação>
_____
_____
_____ Comandos CRIPTA
_____
FIM <identificação>
```

onde '*identificação*' é uma palavra reservada da linguagem. Estas PROCs são procedimentos pré-definidos (procedimentos automatizados), para os quais o usuário deve passar os parâmetros. Dentro destes blocos deverão ser especificados os comandos da Linguagem.

Exemplo de programa CRIPTA

Apresentamos um programa de crítica em CRIPTA, que processa um questionário completo a cada vez, armazenando em memória os registros que o compõem (PROC QUEBRA). O questionário está formado por um único registro de domicílio (tipo 1) e vários registros de pessoa (tipo 3), sendo que a identificação do mesmo está dada pela concatenação das variáveis V10 e V15.

Logo após deve se especificar a PROC CRITICA, para indicar a identificação do relatório, o título e a forma de quebra para impressão do mesmo (V17);

Este programa contém um procedimento especificado em CRIPTA pelo usuário (PROC %...) para gerar a variável CLASSE DE IDADE (W225), a partir dos valores da variável IDADE (V025), quando estivermos processando os registros de pessoas (V100 = 3). Para tanto, o usuário dispõe do processo automatizado PROC GRUPAMENTO;

Em seguida o usuário comanda a incorporação das críticas, através do comando COPIAPLANO, que permite a incorporação, ao programa, das críticas especificadas pelo usuário.

```

PROGRAMA 'EXEMPLO';

    DICIONARIO = BASICO:'XXX',AUXILIAR:'YYYY';

    PROC QUEBRA;
        CHAVE = V10,V15;
    FIM QUEBRA;

    PROC CRITICA;
        IDENTIFICACAO = V10,V15,V100;
        TITULO = 'XXXXXXXXXXXXXXXXXX' +V17;
    FIM CRITICA;

    PROC %AGE_CLASS;

    SE V100 = 3
        ENTAO PROC GRUPAMENTO (W225,V025);
                1 5,
                2 9,
                3 99;

        FIM GRUPAMENTO;

    FIM SE;
FIM %AGE_CLASS;

PROC %CRITICA;
    SE V100 = 1
        ENTAO COPIAPLANO (E10 A E20,E40,E50);
        SENAO COPIAPLANO (E22 A E30);
    FIM SE;
FIM %CRITICA;

FIM PROGRAMA;

```

Nota:

- » E10, E20, E22, E30, E40 e E50 são identificações das regras de edição especificadas pelo usuário no PCAUT (Fig. 3);
- » 'XXXXXXXXXXXXXXXXXX' é o título do relatório de crítica fornecido pelo usuário;
- » o comando DICIONARIO informa ao sistema o DDNAME dos arquivos que contêm os dicionários da aplicação e/ou a identificação do Dicionário Básico na Base de Metadados.

REFERENCIA	CAUSA/PROCEDIMENTO	O P E S C	EFEITO
*****	'CRITICA DE IMPOSSIBILIDADES PARA DOMICILIO'		
E10	'TIPO DE ENTREVISTA INVALIDO' SE INVALIDO (V106);	V	V106: ERRO;
E18	'TOTAL DE QUADRO INVALIDO' SE V209 != SOMAVAR(V202 A V208);	V	V202, V203, V204, V205, V206, V207, V208, V209; ERRO;
E20	'TIPO DE DOMICILIO INVALIDO' SE V201 != QUALQUERUM(1,2,3,5);	V	V201: ERRO;
*****	'CRITICAS DE IMPOSSIBILIDADES PARA PESSOAS'		
E22	'ANO DE NASCIMENTO NAO ACEITAVEL' SE (V310 = BRANCO OU INTERVALO (V310,=100,=888) OU INTERVALO (V310,=991,=998);	V	V310: ERRO;
E26	'VALOR APOSENTADORIA NAO ACEITAVEL' SE V578 = 0;	V	V578: ERRO;
E30	'RECEBE APOSENTADORIA INVALIDO' SE INVALIDO (V5280);	V	V5280: ERRO;
*****	'CRITICAS ENTRE REGISTROS DE DOMICILIO E PESSOAS'		
E40	'TOT. PES. MENOR QUE NUM. PES. COM MAIS 9 ANOS OU DIF. NUM. PES.' W301 = MAIORVALDR(V301 PARA (V100 = 3)); SE V107 != ' ' E (V107<V108 OU V107 DIF W301);	V	V107, V108, V301; ERRO;
E50	'TOT. PES. COM MAIS 9 ANOS DIF. NUM. PES. COM MAIS 9 ANOS' W501 = FREQUENCIA(V100 = 3 E V805 > 9 E (V501 = BRANCO)); SE (V108 != ' ' E V108 DIF W501);	V	V108: ERRO;

- Fig. 3 -

Alternativas Futuras

É possível incrementar o sistema com módulos que realizem outras fases do Ciclo de Vida da Pesquisa, de forma tal a promover um Ambiente Integrado de Ferramentas Automatizadas para geração de Sistemas de Informática para Processamento de Dados Estatísticos.

Para tanto, visualizamos:

- ✓ Interação com sistemas de crítica e correção de dados que implementam outras metodologias (DIA) (já implementado);
- ✓ Facilidades para geração automática de Sistemas de Codificação Assistida (sendo implementado);
- ✓ Facilidades para geração automática de Sistemas de Controle de Processamento (sendo implementado);
- ✓ Portabilidade do sistema, visando a geração das aplicações em distintas plataformas, estações de trabalho (UNIX, DOS) e Sistema Central IBM;
- ✓ Substituição da linguagem de especificação de programas por diálogos nos quais os usuários definam as suas necessidades;
- ✓ Incorporação de facilidades para geração de aplicações voltadas à Disseminação.

Bibliografia

- (1) CRIPTA - Proposta de Desenvolvimento. DI/SUSIS/DISME, Novembro, 1986. (disponível na DI/DETEC/DISIA).
- (2) CRIPTA - Sistema automatizado de Plano de Crítica, Imputação Automática e Especificação de Procedimentos Dedicados. DI/GESID, Abril, 1987.
- (3) BARBOSA, D.M.R. & HANONO, R.M. - Estudos das Ferramentas para Apuração de Dados. Revista Brasileira de Estatística, Rio de Janeiro, 49(191), Jan/Jun, 1988, 85-100.
- (4) CRIPTA - Especificação Lógica. DI/GESIE (disponível na DI/DETEC/DISIA).
- (5) BARBOSA, D.M.R. - CRIPTA - Arquitetura do Sistema. (disponível na DI/DETEC/DISIA).
- (6) BARBOSA, D.M.R. - CRIPTA - Estrutura da Gramática. (disponível na DI/DETEC/DISIA).
- (7) CRIPTA - Avaliação do Sistema durante o Beta Teste. (disponível na DI/DETEC/DISIA).
- (8) CABRAL, M.S.S. - Um Esquema para Apuração de Pesquisas Estatísticas do IBGE. Artigo Técnico A011/90, IBGE.
- (9) CABRAL, M.S.S. - O Esquema Genérico para a Apuração das Pesquisas do Statistics Canada. Artigo Técnico A005/90, IBGE.
- (10) CRIPTA - Manual da Linguagem. IBGE/DI/Biblioteca.
- (11) CRIPTA - Manual do Dicionário da Aplicação. IBGE/DI/Biblioteca.
- (12) CRIPTA - Manual do Plano de Crítica Automatizado. IBGE/DI/Biblioteca.
- (13) CRIPTA - Manual de Estudo de Casos. IBGE/DI/Biblioteca.
- (14) GUEDES, A.P. - Banco de Metadados. Novembro, 1989, DEBAD/GEBAD - IBGE/DI/Biblioteca.
- (15) HANONO, R.M. & BARBOSA, D.M.R. - Uma Experiência de Desenvolvimento de Sistema Gerador de Aplicações utilizando Recursos de Engenharia de Software. Artigo Técnico A016/90, IBGE.
- (16) RUBIO, E.G. & CRIADO, I.V. - Sistema DIA - Sistema de Detección e Implementación de Errores para Datos Cualitativos. Descripción del Sistema (Vol. I). INE/Madrid (España), 1988.
- (17) SAS 'Statistical Analysis System' User Guide. SAS Institute Inc.
- (18) ATLAS - Manual do Usuário. Manual do Sistema. IBGE/DI/Biblioteca.
- (19) TSGBD/X - Manual de Projeto de Banco de Dados V.1.0. Tecnocoop Sistemas.
- (20) IMPS - Integrated Microcomputing Programing System, V.5, U.S. Census Bureau.
- (21) REDATAM - Retrieval of Census Data for Small Areas in Microcomputers, Latin America Demographic Center (CELADE) of the United Nations Economic Commission for Latin America and Caribbean.
- (22) BRAGA, L.C.A. & BRAZ, R. - Alfa Teste do CRIPTA. Artigo Técnico A010/90, IBGE.