Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
Departamento de Ciências de Computação
SCC0222 – Laboratório de Introdução à Ciência da Computação I

Trabalho 07 (sub): Captcha

Professor: Dr. Rodrigo Fernandes de Mello (mello@icmc.usp.br)
Estagiário PAE: Fábio Henrique Gomes Sikansi (fhenrique@usp.br)
Martha Dais Ferreira (daismf@icmc.usp.br)
Monitor: Lucas Parras (parraslucas@gmail.com)
Loys Gibertoni (loys.gibertoni@usp.br)
Colaborador: Felipe Simões Lage Gomes Duarte (fgduarte@icmc.usp.br)

1 Objetivo do Trabalho

Você deverá implementar um sistema capaz de quebrar um captcha no formato PGM.

2 Imagem PGM

A designação de formato de imagem PBM (Portable Bitmap) engloba três formatos de imagem para imagens a preto e branco, em escala de tons cinzentos e a cores, todos eles sem compressão e que apresentam uma estrutura comum. Estes três tipos de formato de imagens são:

- PBM (Portable BitMap) - imagens a preto e branco (sem tons de cinzento)
- PGM (Portable GrayMap) - imagens em tons de cinzento
- PPM (Portable PixMap) - imagens a cores

A definição original destes formatos teve em vista permitir a transmissão de imagens por meio de correio eletrônico que até então não permitia a transmissão de ficheiros anexados, binários ou não. A definição do formato foi mais tarde modificada para permitir a representação binária dos conteúdos das imagens.

Os formatos de imagem PGM são constituídos pelos seguintes campos:

```
<magic number>
<Altura da imagem ‘‘m’’> <Largura da imagem ‘‘n’’>
<Valor máximo dos tons de cinza>
<valor do pixel (0,0)> ... <valor do pixel (0,n)>
  .          .              .
  .             .           .
  .                .        .
<valor do pixel (m,0)> ... <valor do pixel (m,n)>
```

O Identificador do tipo de formato (designado por “magic number”), é determinado de acordo com Tipo, ASCII ou Binário

Tipo	ASCII	Binário
PBM	P1	P4
PGM	P2	P5
PPM	P3	P6

Em particular, no nosso caso, iremos trabalhar com imagens PGM do tipo ASCII, ou seja, “P2”. Assim, um exemplo de imagem PGM que o programa deve ser capaz de processar é:

```
P2
24 7
15
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  3  3  3  3  0  0  7  7  7  7  0  0 11 11 11 11  0  0 15 15 15 15  0
0  3  0  0  0  0  0  7  0  0  0  0  0 11  0  0  0  0  0 15  0  0 15  0
0  3  3  3  0  0  0  7  7  7  0  0  0 11 11 11  0  0  0 15 15 15 15  0
0  3  0  0  0  0  0  7  0  0  0  0  0 11  0  0  0  0  0 15  0  0  0  0
0  3  0  0  0  0  0  7  7  7  7  0  0 11 11 11 11  0  0 15  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
```

3 Captcha

CAPTCHA é um acrônimo da expressão "Completely Automated Public Turing test to tell Computers and Humans Apart" (teste de Turing público completamente automatizado para diferenciação entre computadores e humanos): um teste de desafio cognitivo, utilizado como ferramenta anti-spam, desenvolvido pioneiramente na universidade de Carnegie-Mellon. Como o teste é administrado por um computador, em contraste ao teste de Turing padrão que é administrado por um ser humano, este teste é na realidade corretamente descrito como um teste de Turing reverso.

Um CAPTCHA usual envolve um computador (um servidor) que pede que um usuário termine um teste. Como os computadores, na teoria, são incapazes de resolver o CAPTCHA, todo usuário que incorpora uma solução correta é presumidamente humano. O termo foi inventado em 2000 por Luis von Ahn, por Manuel Blum, Nicholas J. Hopper (todos da universidade do Carnegie-Mellon), e por John Langford (da IBM).

Um tipo comum de CAPTCHA requer que o usuário identifique as letras de uma imagem distorcida, às vezes com a adição de uma sequência obscurecida das letras ou dos dígitos que apareça na tela.

4 Proposta

O sistema deve ser capaz de ler um captcha que está armazenado em um arquivo do tipo PGM-P2 e escrever no stdout o valor presente na imagem. Para isso, seu programa receberá como entrada somente o nome do arquivo:

```
captcha_01.pgm
```

O arquivo informado é uma image PGM-P2 binaria com ruído do tipo sal e pimenta. Ela descreve um captcha formado apenas por números no formato digital tal como na imagem abaixo:

Figura 1: Exemplo de Captcha

Para auxiliar, está em anexo no run.codes um arquivo zip contendo um pgm de cada número digital, você pode utiliza-los para orientar o desenvolvimento de seu trabalho.

O seu programa deverá imprimir o resultado final tal como o exemplo abaixo:

```
123456
```

5 IMPORTANTE

- É obrigatório o uso de alocação dinâmica para qualquer vetor/matriz utilizado em seu programa.
- Crie quantas funções achar necessário, a modularização do sistema será levado em consideração no processo de correção.
- A metodologia para solução do problema está em aberto, pesquisa e resolva como achar melhor e documente no começo do código qual a abordagem escolhida para resolver.