



# SCARLET – Desenvolvimento de uma solução de compressão aplicada a dados textuais estruturados em SVG

Anderson Lemos da Silva

Orientador: Arthur Rodrigues Araruna

# Introdução

- Necessidade da informação
- Sistemas computacionais
- Problemas de transmissão e armazenamento
  - ✓ Dados estruturados
  - ✓ Compressão de dados
- Imagens na web

# Fundamentação Teórica

- Dados estruturados
  - ✓ “Dados estruturados são dados dispostos em representações rígidas, sujeitas a regras e a restrições impostas pelo esquema que os criou.” (ALMEIDA, 2002).

# Fundamentação Teórica

- Dados estruturados
  - ✓ XML
  - ✓ SVG

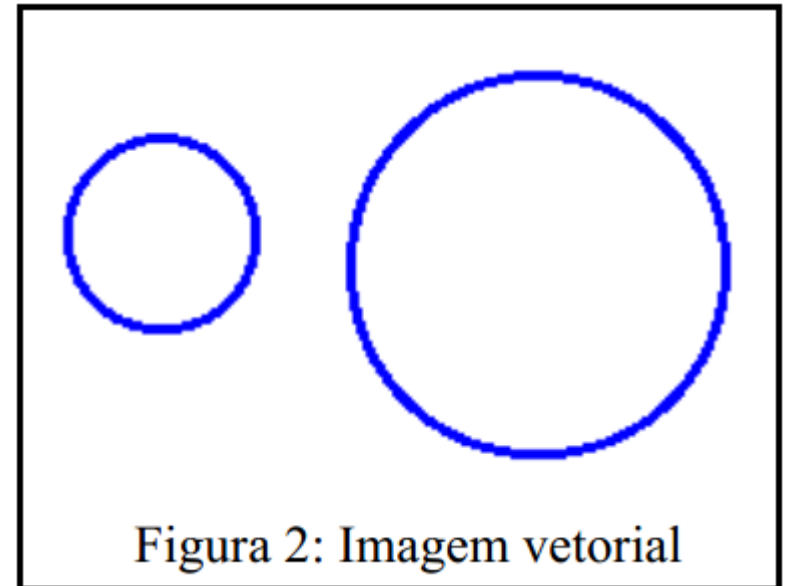
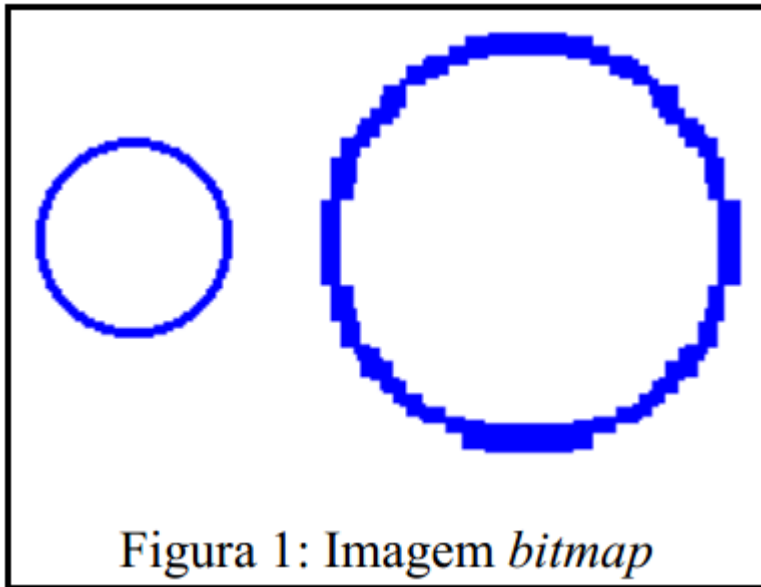
# Fundamentação Teórica

- XML

```
<document>
  <report>
    <author>Video database</author >
    <date>June 12, 2000</date>
  </report >
  <paper>
    <title>XML query data model</title>
    <author>Don Robie</author>
    <source>W3C, June 2000</source>
  </paper>
</document>
```

# Fundamentação Teórica

- SVG
  - ✓ Imagens Bitmap x Imagens Vetoriais



# Fundamentação Teórica

- Compressão de dados
  - ✓ “Compressão de dados é o processo de converter um conjunto de entrada de dados – dados originais – em um conjunto de dados de saída – dados comprimidos –, que possui um menor tamanho comparado aos dados originais” (TEIXEIRA, 2011)
  - ✓ Descompressão

# Fundamentação Teórica

- Técnicas de compressão de dados
  - ✓ Binário x Texto
  - ✓ Genéricas x Dedicadas



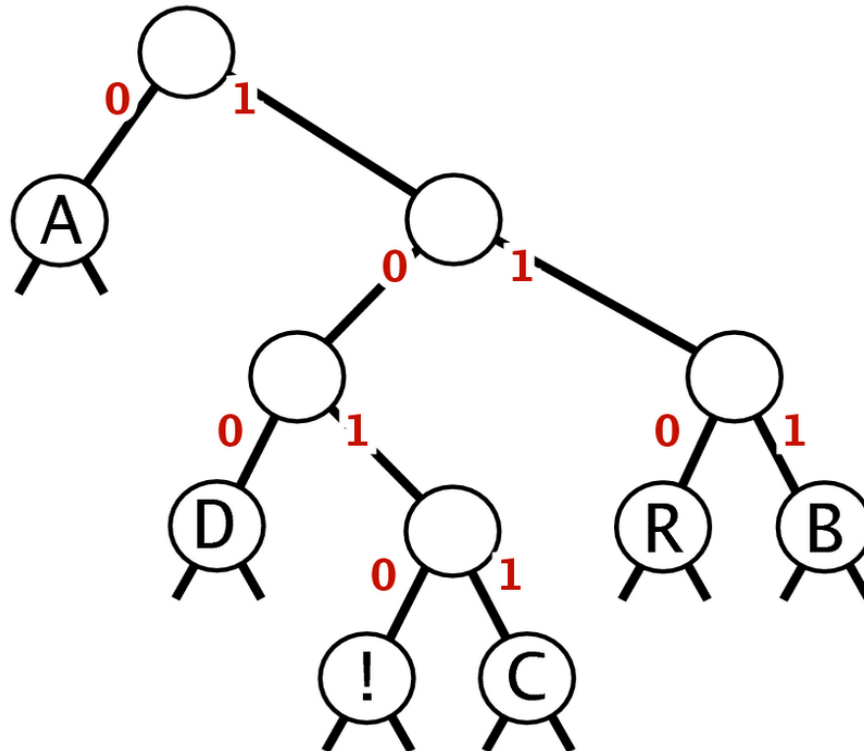
# Fundamentação Teórica

- Codificação de Huffman

**Tabela  
palavra-chave**

Chave	Valor
!	1010
A	0
B	111
C	1011
D	100
R	110

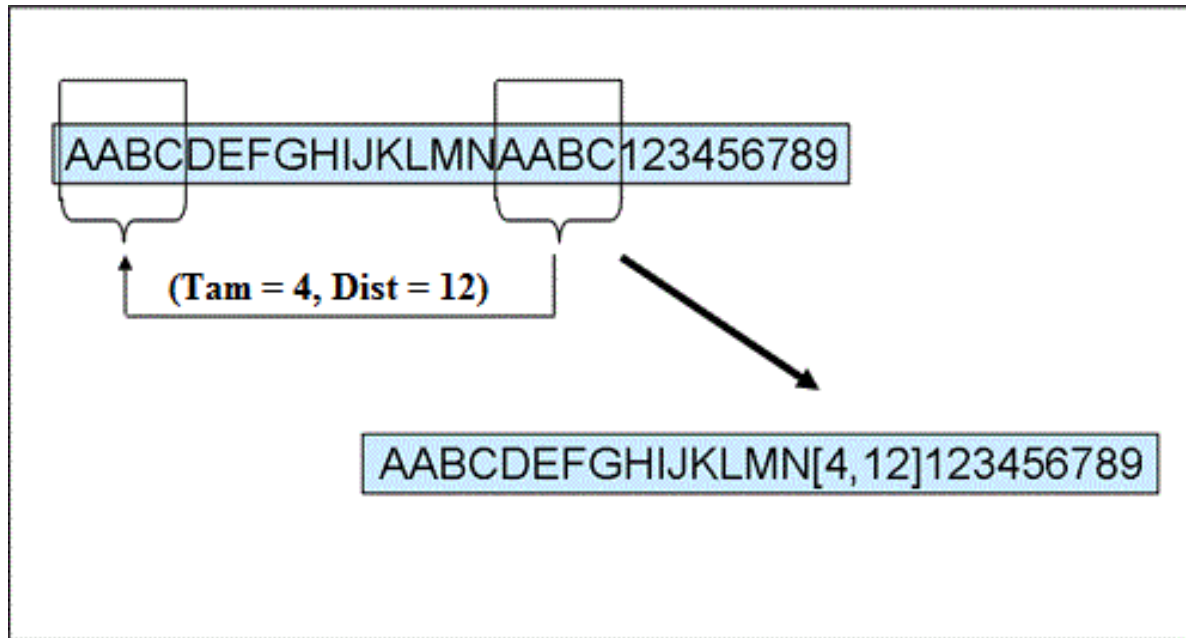
**Representação em árvore**



**ABRACADABRA!**

# Fundamentação Teórica

- LZ77



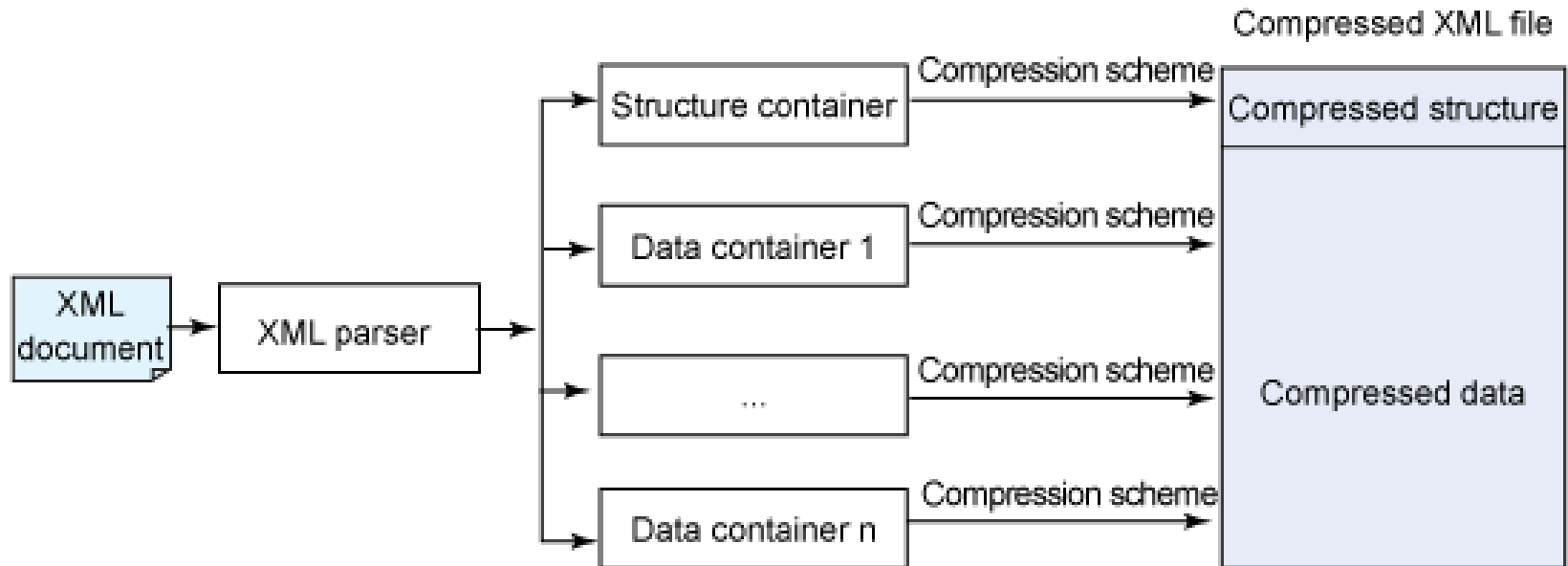
- Zlib e Gzip

# Compressão de dados XML

- Propósito geral
- XML-conscious
  - ✓ Esquema informado
  - ✓ Esquema não informado
- Técnicas pesquisáveis

# Compressão de dados XML

- XMill



# Trabalhos relacionados

- **SAKR, S. Xml compression techniques: A survey and comparison.** Journal of Computer and System Sciences, Elsevier, v. 75, n. 5, p. 303–322, 2009.

- ✓ Levantamento do estado da arte
- ✓ Comparação entre nove técnicas

# Trabalhos relacionados

- **TEIXEIRA, Márlon A. C. Novas Abordagens para Compressão de Documentos XML.** 2011. Dissertação (Mestrado em Engenharia Elétrica) - Departamento de Comunicações, Universidade Estadual de Campinas, Campinas. 2011.
  - ✓ Duas novas abordagens de compressão para XML
  - ✓ Comparativo com abordagens já existentes

# Trabalhos relacionados

• **KHEIRKHAHZADEH, A. D. On The Performance of Markup Language Compression.** Tese (Doutorado)  
— University OF West London, 2015.

- ✓ Estudo em compressores de comprimento fixo aplicada a técnicas de propósito geral.
- ✓ Desenvolvimento de uma nova técnica e comparativo.

# Metodologia e Desenvolvimento

- Revisão bibliográfica da área de compressão de dados
- Eleição dos principais métodos de compressão aplicáveis a dados SVG
- Análise das técnicas e tecnologias
  - ✓ C++ e QT
- Montagem dos repositórios de testes
  - ✓ Repositório 1: 1000 arquivos
  - ✓ Repositório 2: 3000 arquivos



# Metodologia e Desenvolvimento

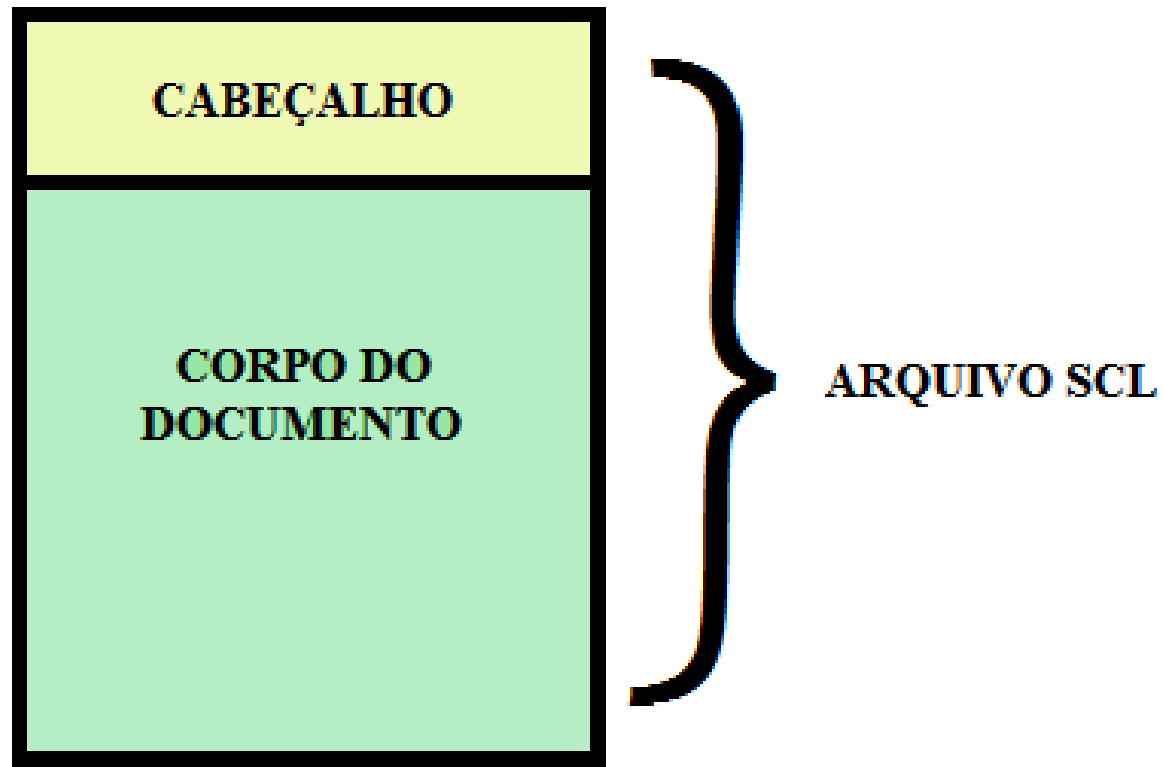
- Análise estatística
  - ✓ Nomes de tags e atributos
    - 241 tags
    - 477 atributos
    - 7 iguais
  - ✓ Conteúdo das tags e atributos
    - Conteúdo numérico
    - Conteúdo textual

# Metodologia e Desenvolvimento

- Desenvolvimento da nossa solução - SCARLET
  - ✓ XML-conscious
    - Esquema não informado
  - ✓ Codificação de Huffman
    - Tags
    - Atributos
    - Conteúdo numérico
    - Conteúdo Textual

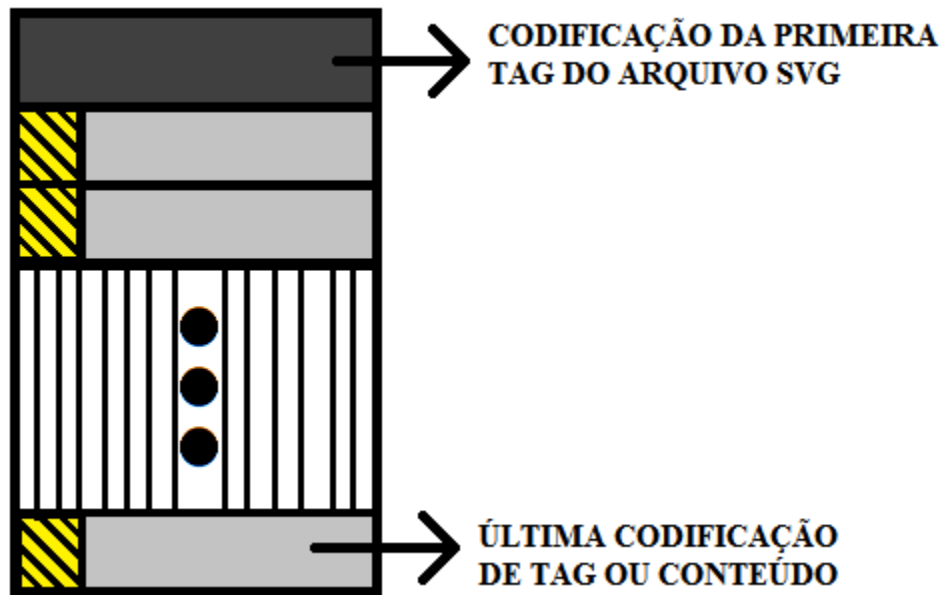
# Metodologia e Desenvolvimento

- Padrão do arquivo compactado



# Metodologia e Desenvolvimento

- Corpo do documento



## LEGENDA:



CODIFICAÇÃO DE TAG



SEPARADOR



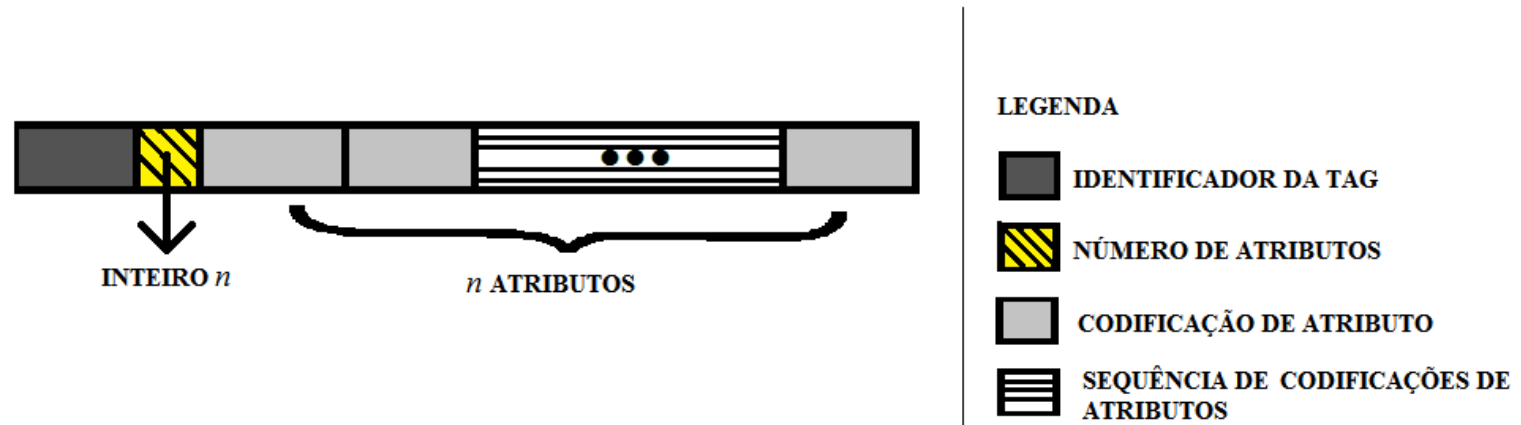
CODIFICAÇÃO DE TAG OU  
CONTEÚDO



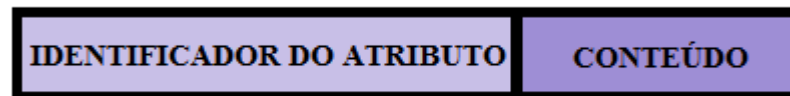
SEQUÊNCIA DE SEPARADORES E  
TAGS/CONTEÚDOS DO ARQUIVO

# Metodologia e Desenvolvimento

- Codificação de Tag

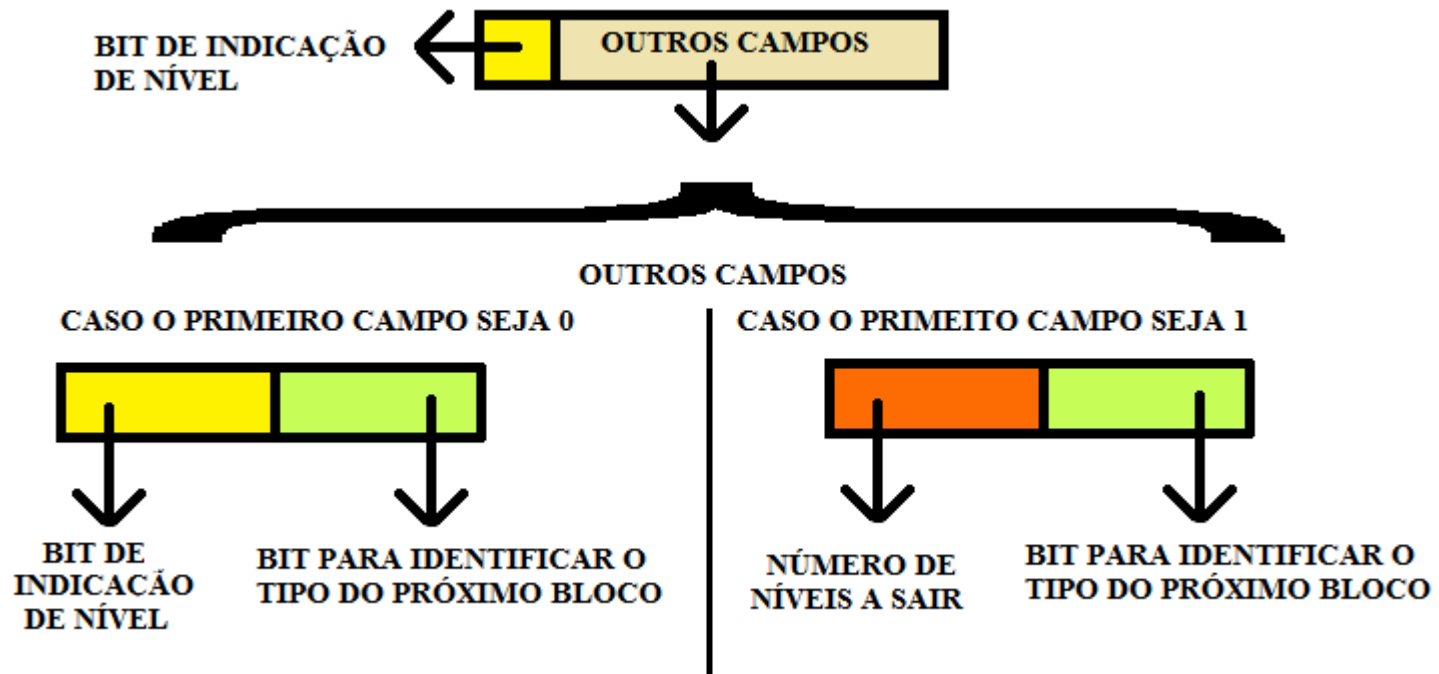


- Codificação de atributo



# Metodologia e Desenvolvimento

- Codificação de separador



# Metodologia e Desenvolvimento

- Implementação das técnicas
- Execução dos métodos
  - ✓ Cinco execuções
  - ✓ Informações coletadas:
    - Tamanho arquivo comprimido
    - Taxa de compressão
    - Tempo de execução
    - Tamanho do arquivo descomprimido
    - Tamanho Teórico Mínimo
- Eleição do melhor

# Resultados

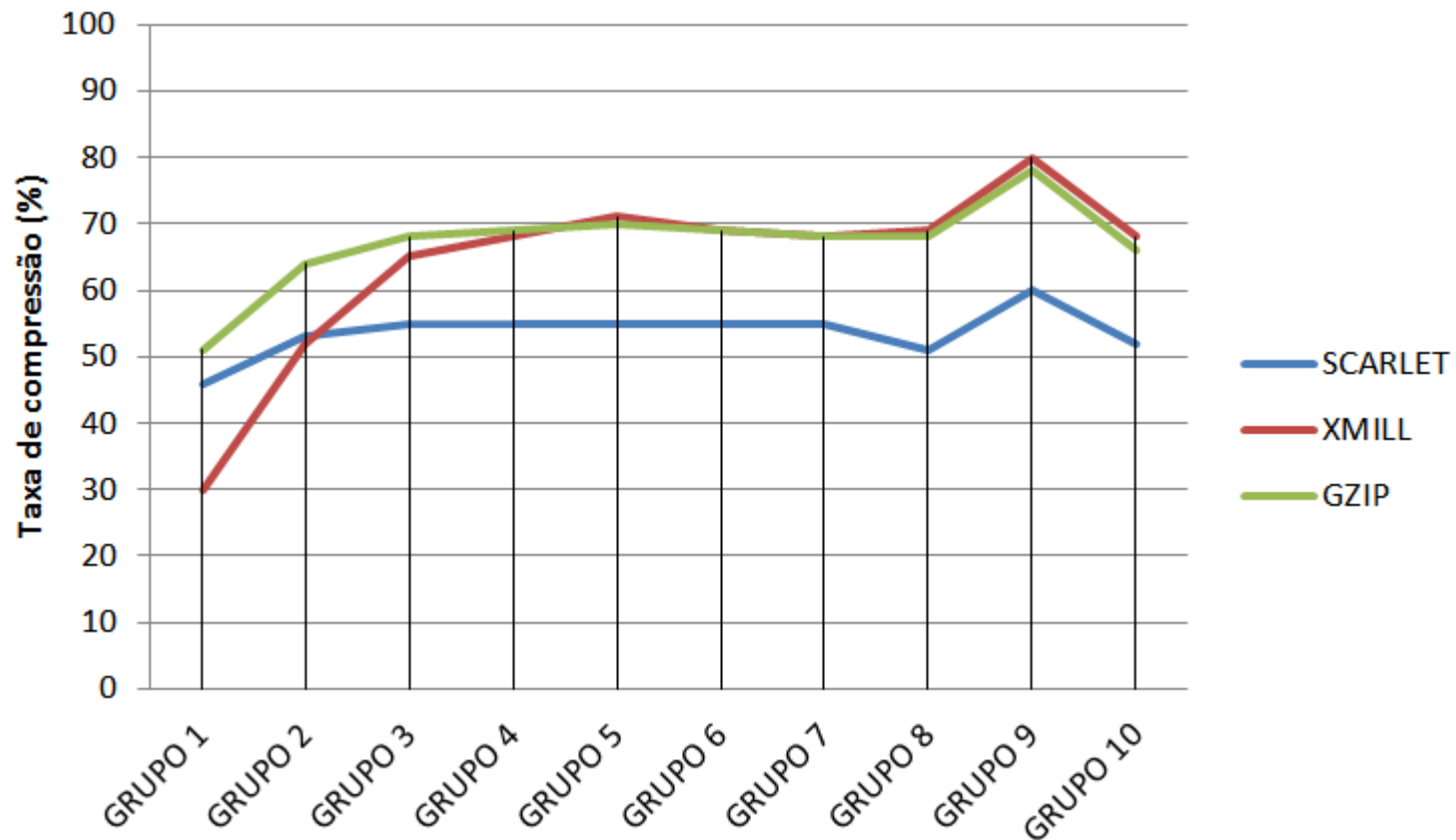
- Resultados em grupos

<b>Grupo</b>	<b>Faixa de tamanho dos arquivos</b>	<b>Quantidade de arquivos</b>
Grupo 1	<i>214 Bytes - 5 KBytes</i>	491
Grupo 2	<i>5 KBytes - 10 KBytes</i>	752
Grupo 3	<i>10 KBytes - 20 KBytes</i>	211
Grupo 4	<i>20 KBytes - 50 KBytes</i>	258
Grupo 5	<i>50 KBytes - 100 KBytes</i>	243
Grupo 6	<i>100 KBytes - 200 KBytes</i>	315
Grupo 7	<i>200 KBytes - 500 KBytes</i>	292
Grupo 8	<i>5 KBytes - 1 MByte</i>	111
Grupo 9	<i>1 MByte - 5 MBytes</i>	294
Grupo 10	<i>5 MBytes - 27 MBytes</i>	33



# Resultados

- Taxa de compressão



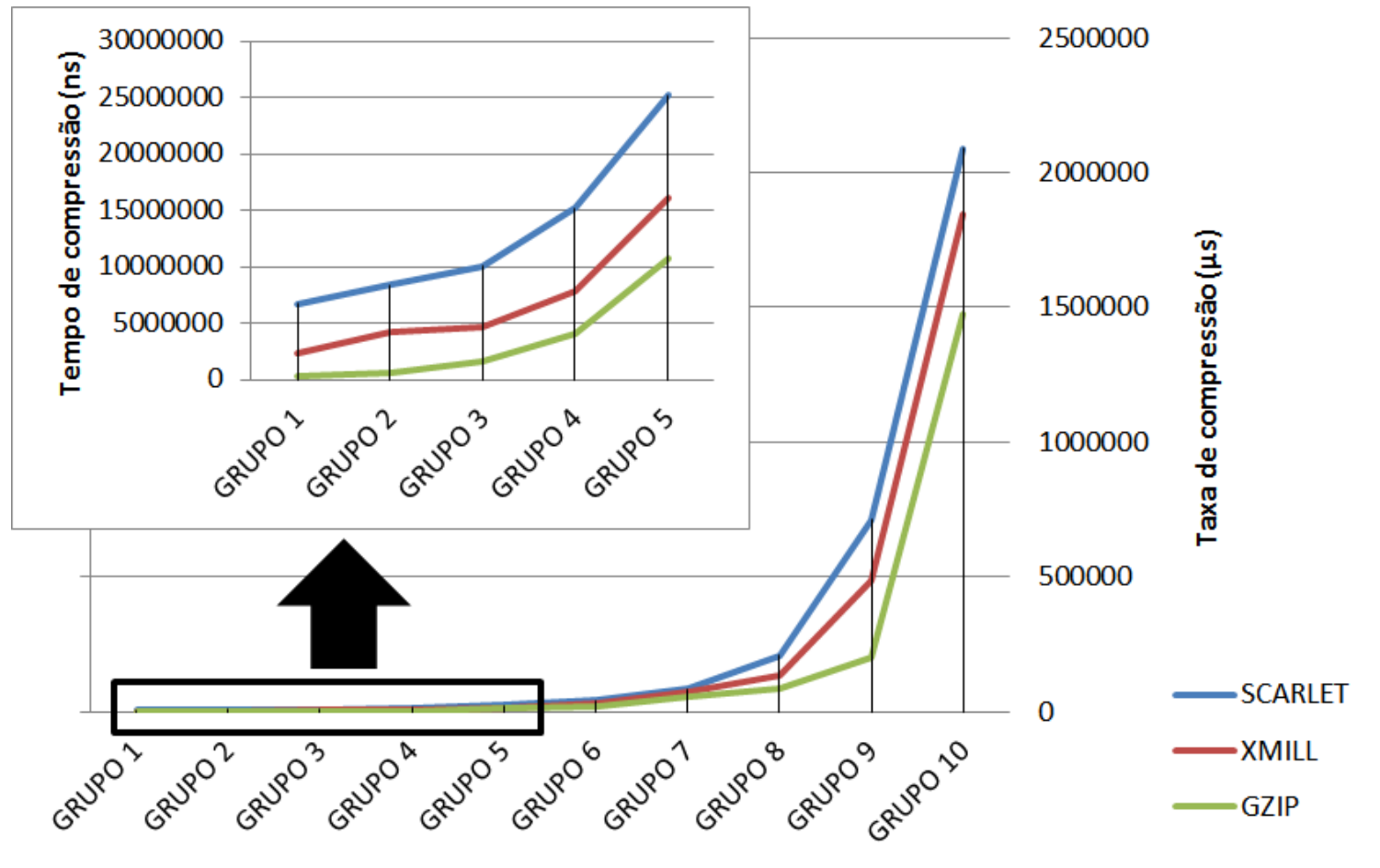
# Resultados

- Taxa de compressão – Maiores arquivos de cada grupo



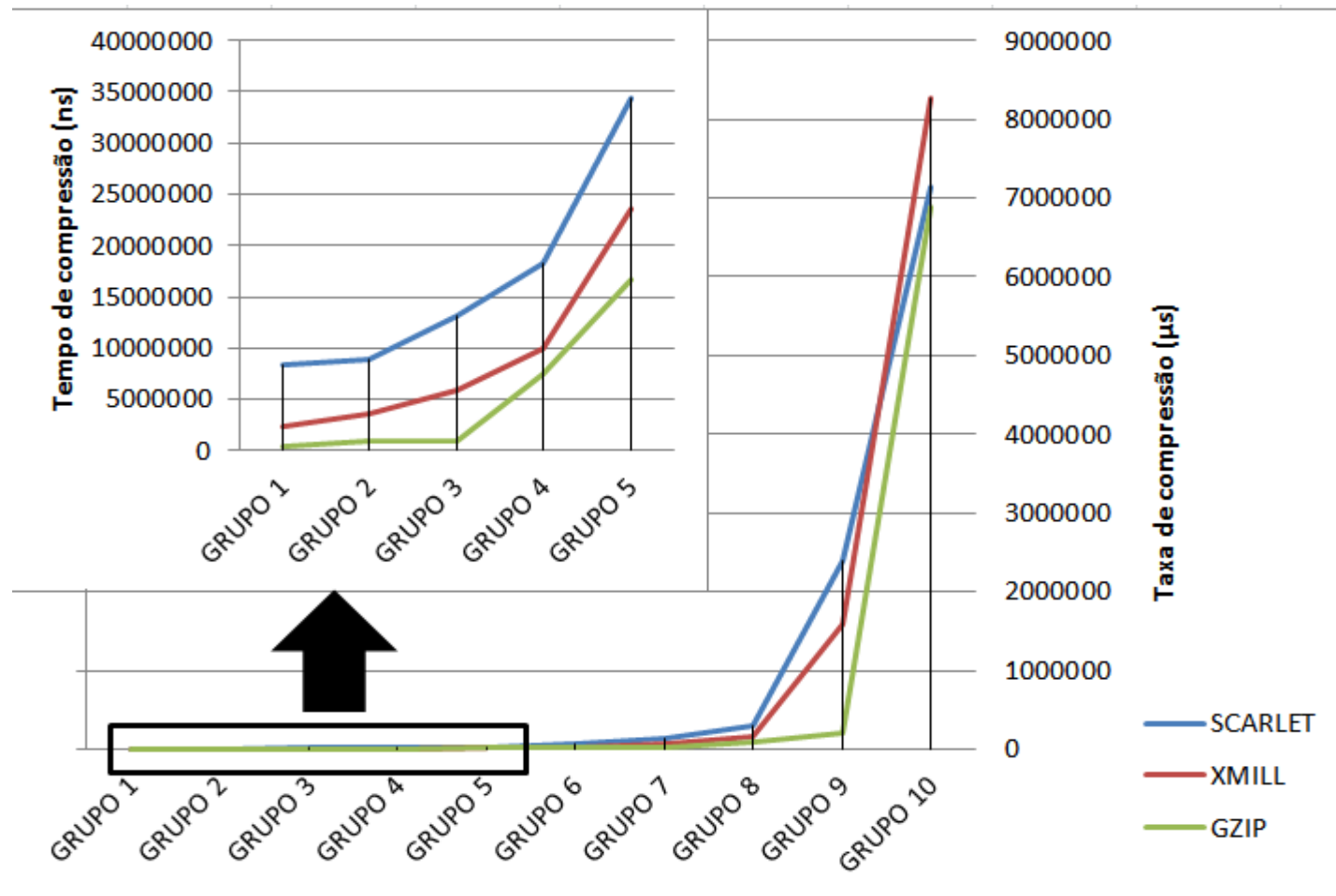
# Resultados

- Tempo de compressão



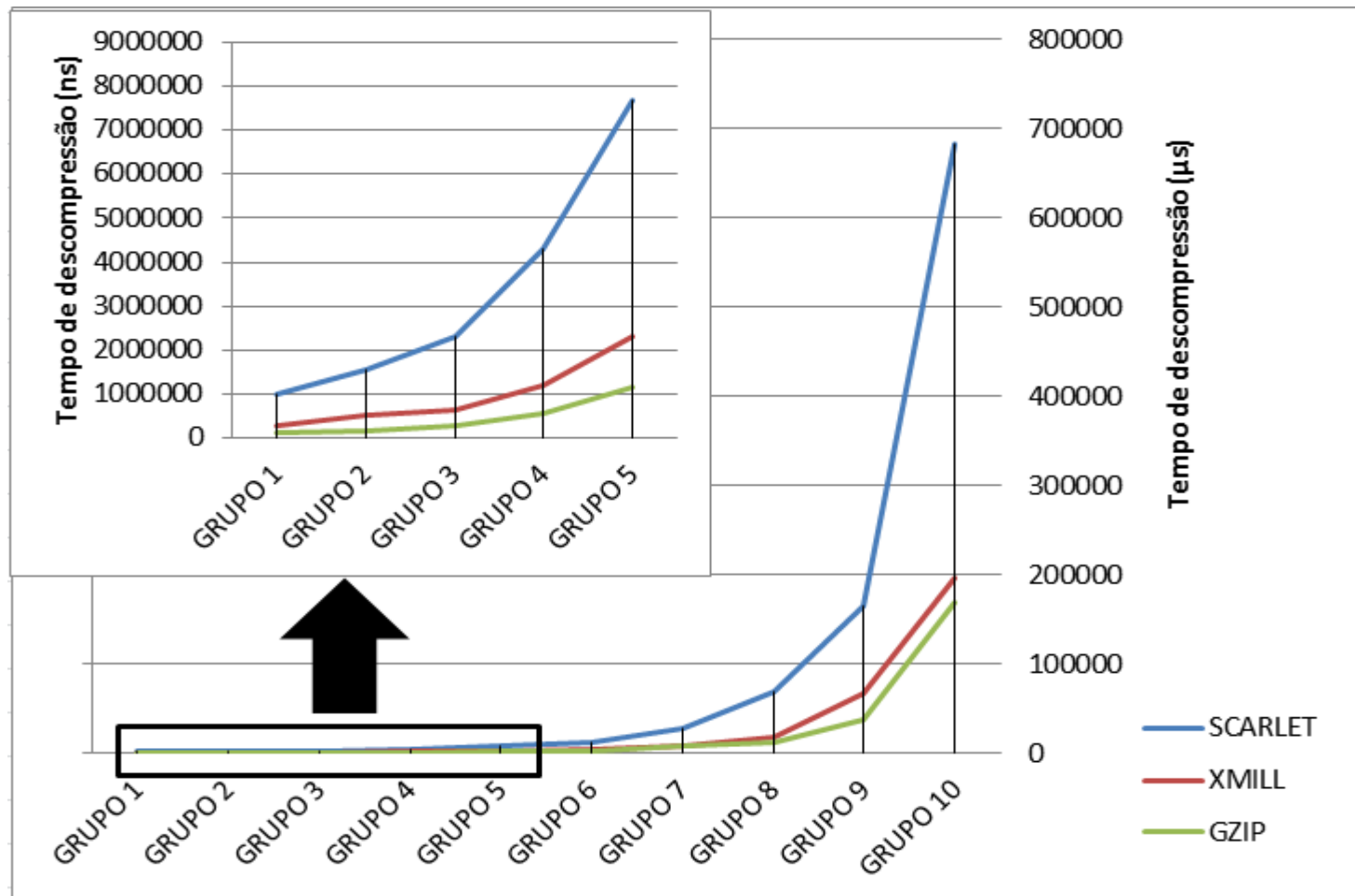
# Resultados

- Tempo de compressão – Maiores arquivos de cada grupo



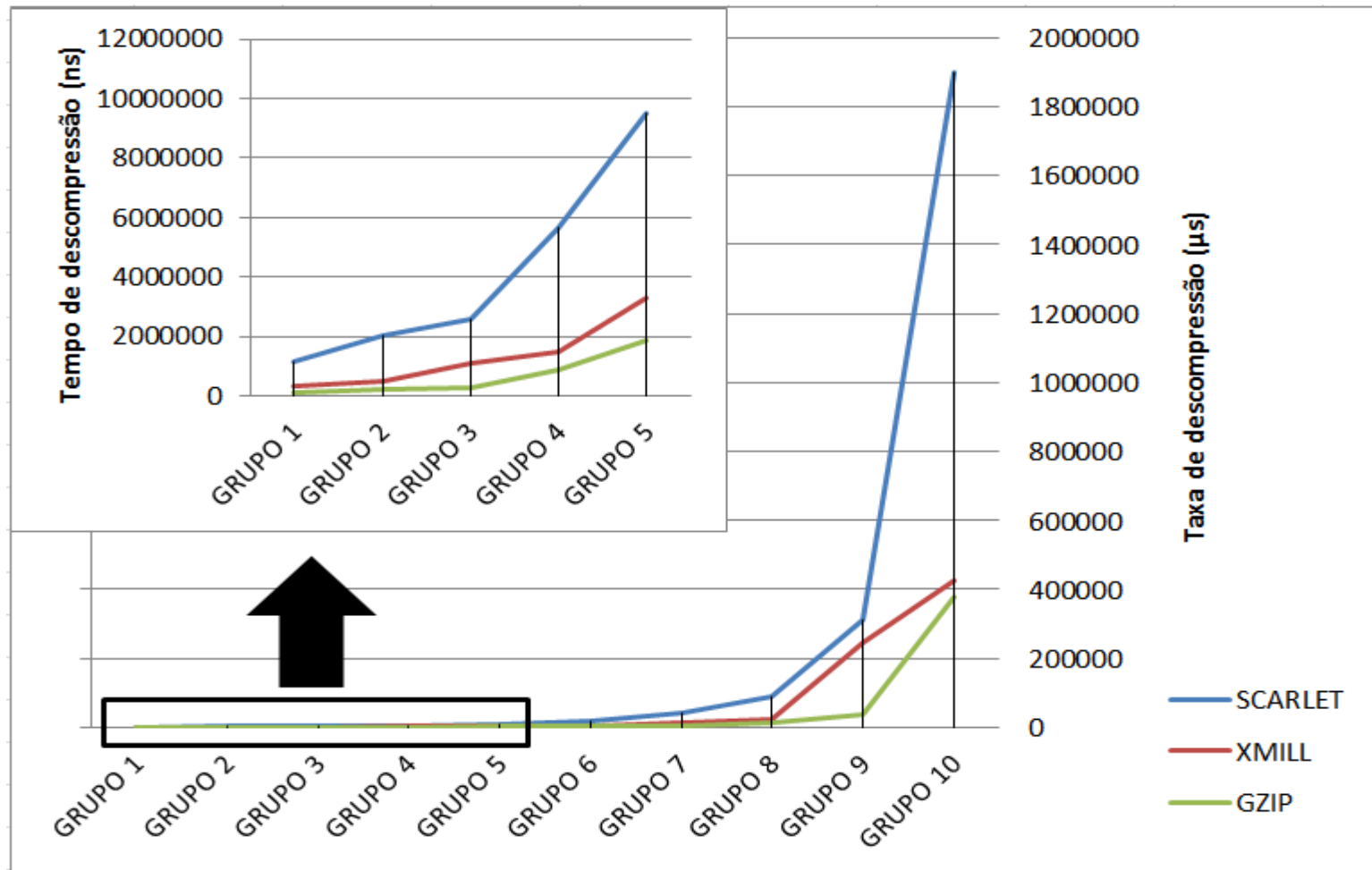
# Resultados

- Tempo de descompressão



# Resultados

- Tempo de descompressão – Maiores arquivos de cada grupo



# Considerações Finais

- Gzip
- Bom desempenho do SCARLET
- Trabalho Futuros

Obrigado!