Análise de ocorrência de falsos positivos em recuperação de dados formatados

Rubens K. P. Silva, Islan A. Bezerra, Sidney M. L. de Lima, Sérgio M. M. Fernandes.

Departamento de Engenharia da Computação - Universidade de Pernambuco, (UPE)

Departamento de Eletrônica e Sistemas -- Universidade Federal de Pernambuco (UFPE)

Motivação

- Exclusão de arquivos para ocultar evidências que apontem autoria de ilícitos.

Objetivo

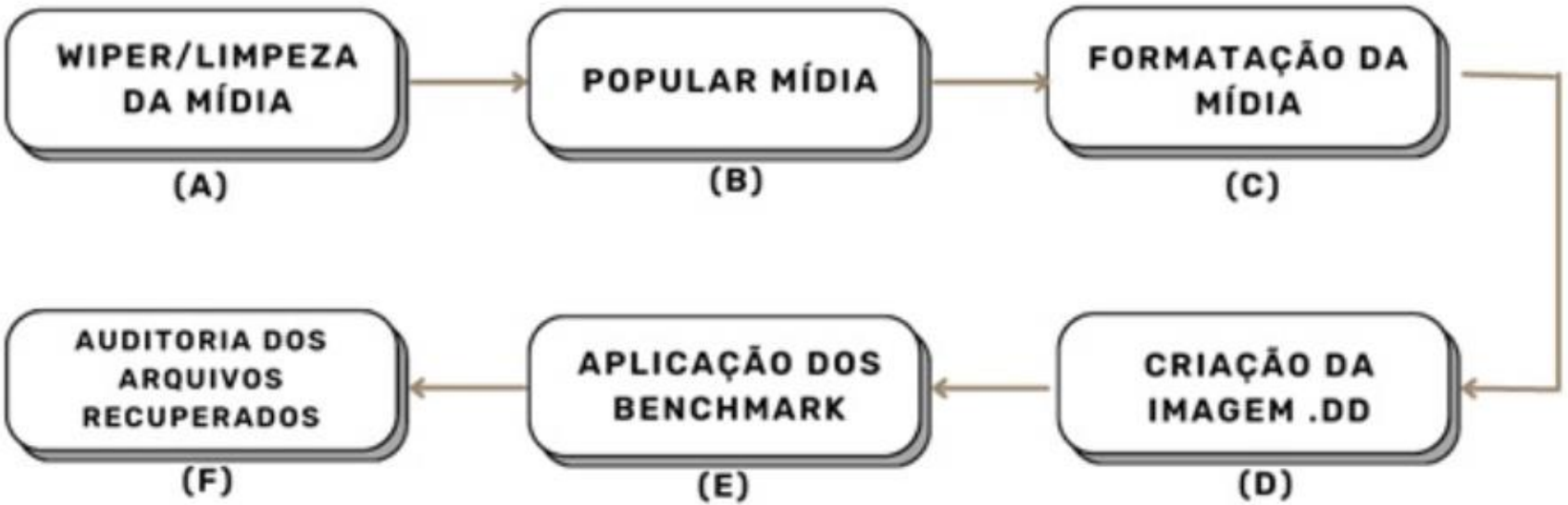
- Analisar recuperação de dados formatados de diferentes softwares, com foco na identificação de falsos positivos.

- Pesquisa experimental.

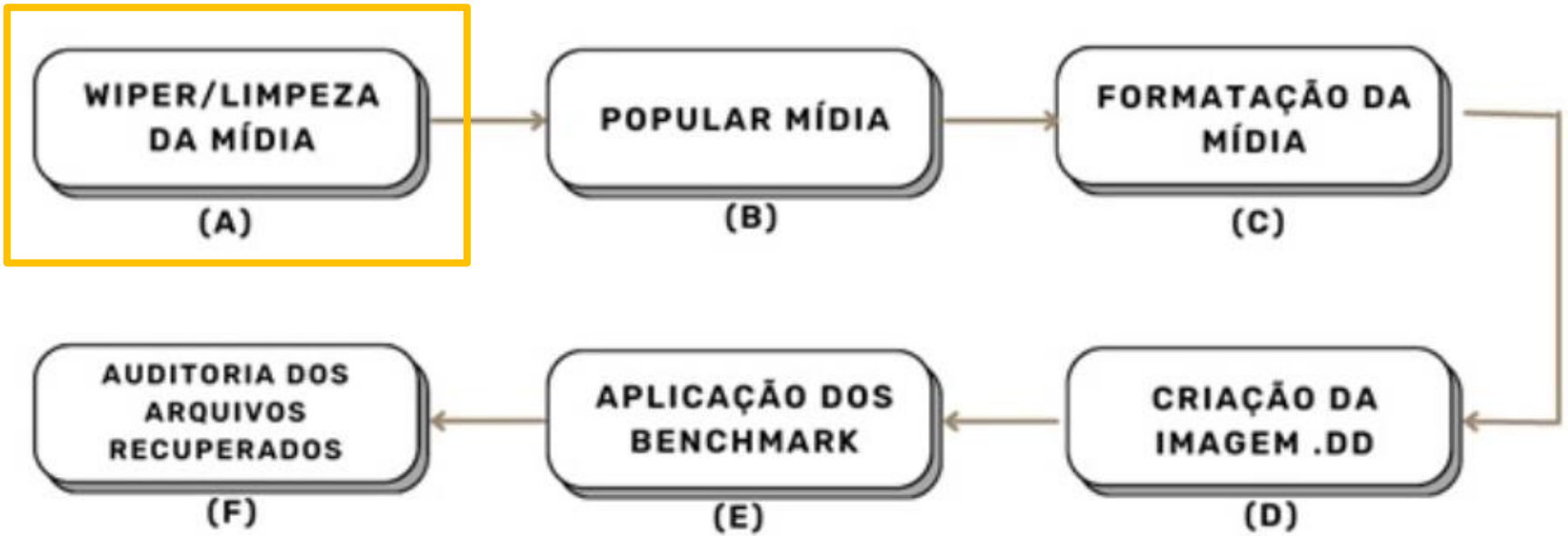
Trabalhos Relacionados

- Pereira et al. 2019 – RSL evidenciando ausência de trabalhos que tratem de falsos positivo.
- Nurhayati 2017 - Proposta de método experimental.

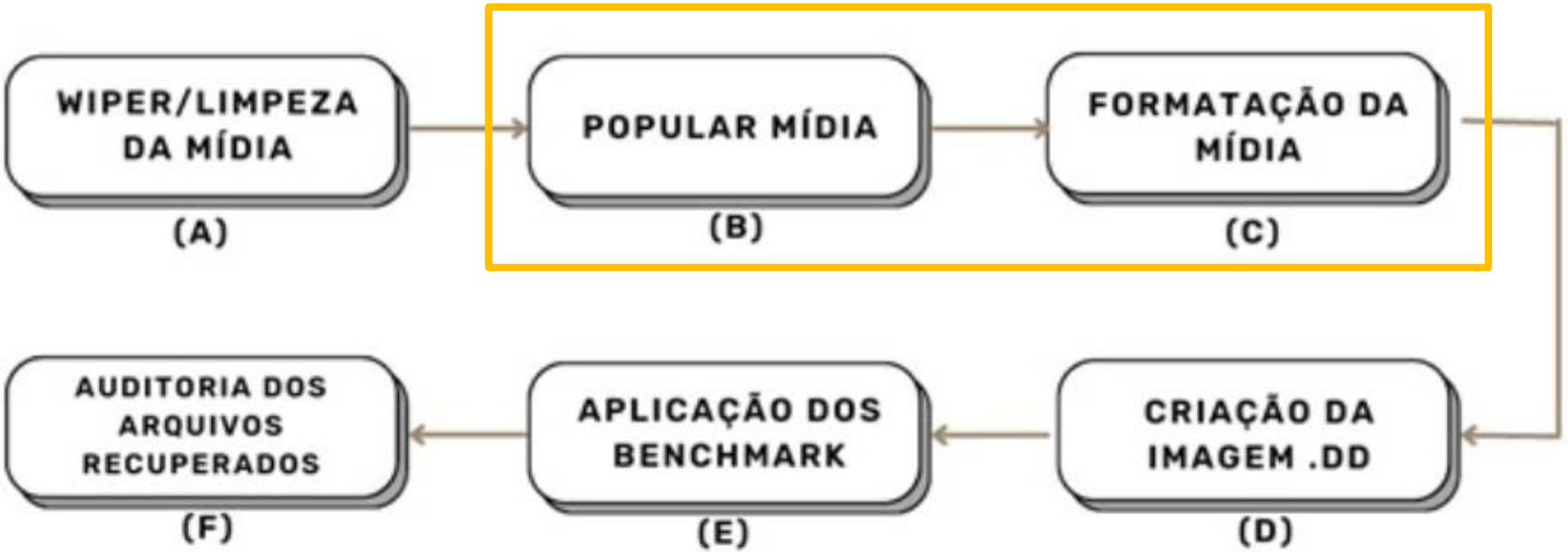
Método


WIPER/LIMPEZA DA MÍDIA
(A)
POPULAR MÍDIA
(B)
FORMATAÇÃO DA MÍDIA
(C)
CRIAÇÃO DA IMAGEM .DD
(D)
APLICAÇÃO DOS BENCHMARK
(E)
AUDITORIA DOS ARQUIVOS RECUPERADOS
(F)

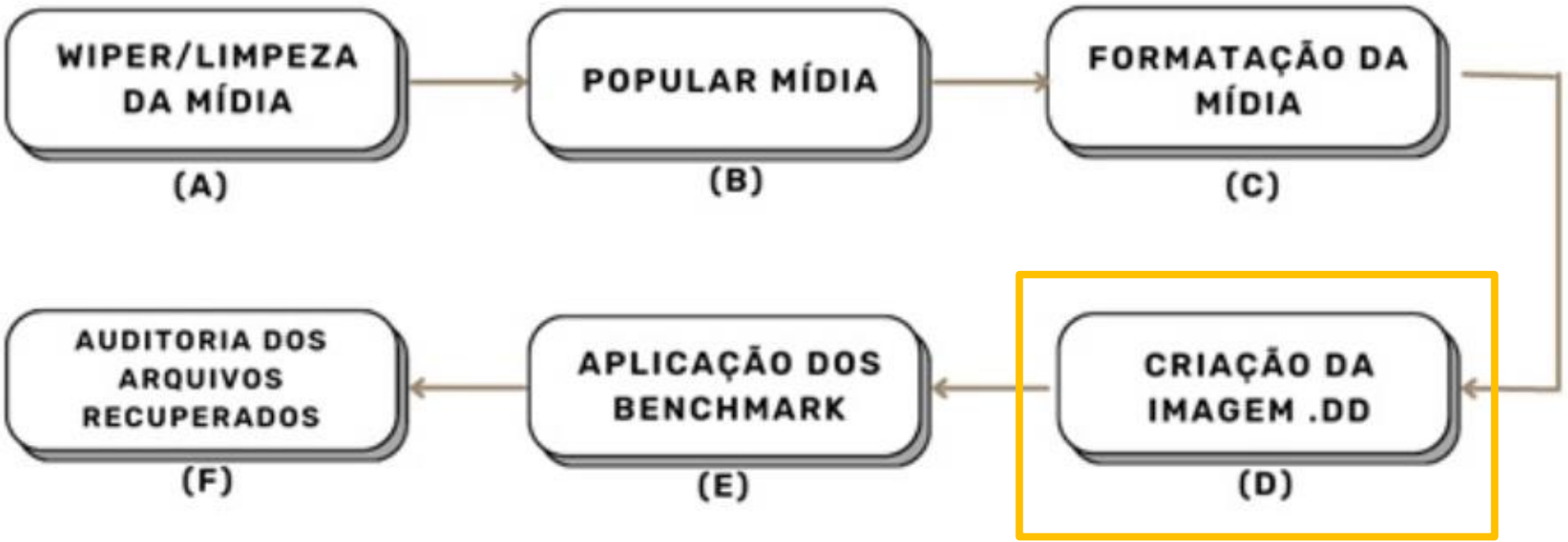
Método


WIPER/LIMPEZA DA MÍDIA
(A)
POPULAR MÍDIA
(B)
FORMATAÇÃO DA MÍDIA
(C)
CRIAÇÃO DA IMAGEM .DD
(D)
APLICAÇÃO DOS BENCHMARK
(E)
AUDITORIA DOS ARQUIVOS RECUPERADOS
(F)

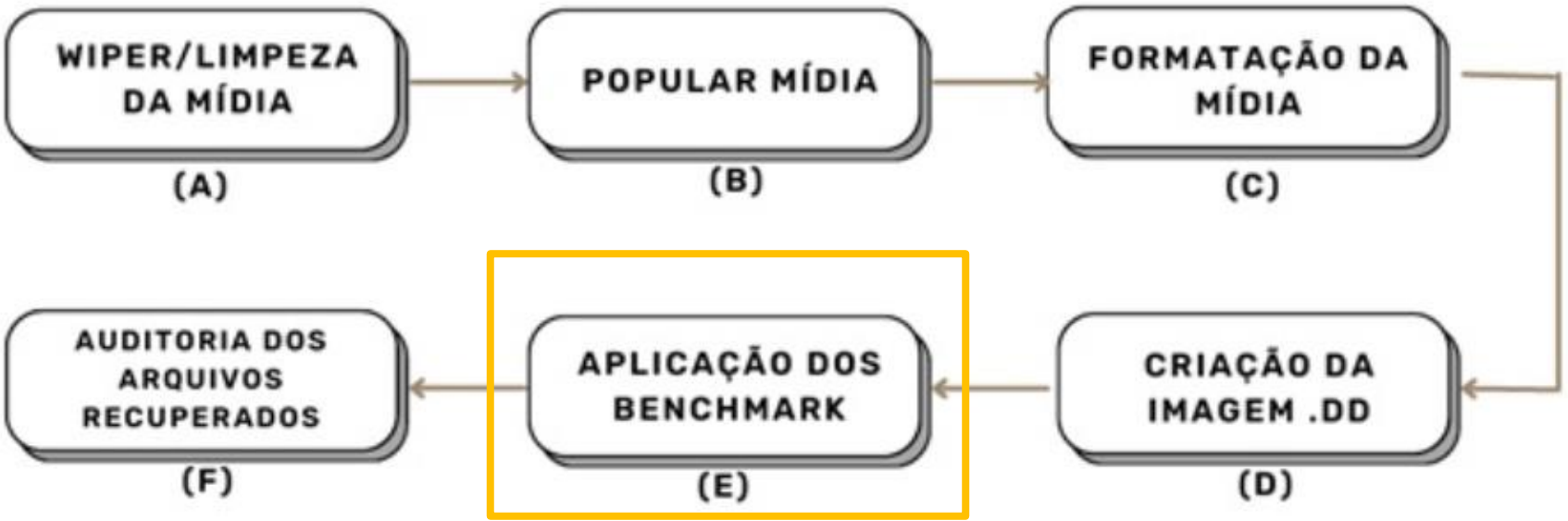
Método


WIPER/LIMPEZA DA MÍDIA
(A)
POPULAR MÍDIA
(B)
FORMATAÇÃO DA MÍDIA
(C)
CRIAÇÃO DA IMAGEM .DD
(D)
APLICAÇÃO DOS BENCHMARK
(E)
AUDITORIA DOS ARQUIVOS RECUPERADOS
(F)

Método


WIPER/LIMPEZA DA MÍDIA
(A)
POPULAR MÍDIA
(B)
FORMATAÇÃO DA MÍDIA
(C)
CRIAÇÃO DA IMAGEM .DD
(D)
APLICAÇÃO DOS BENCHMARK
(E)
AUDITORIA DOS ARQUIVOS RECUPERADOS
(F)

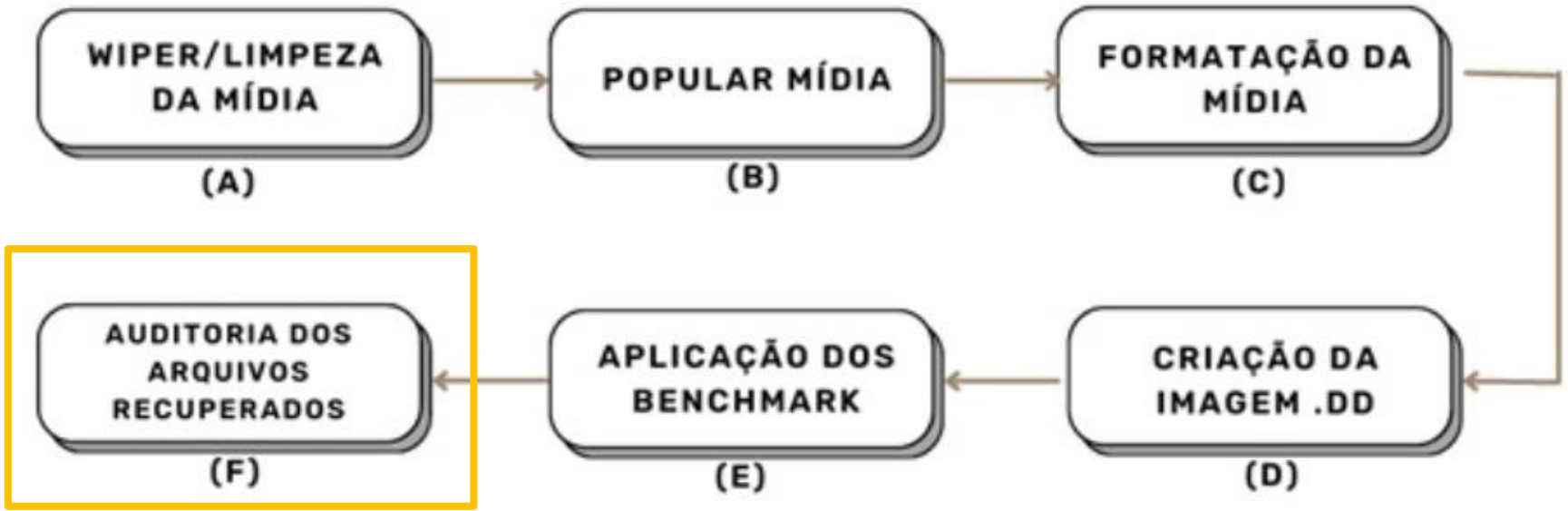
Método


WIPER/LIMPEZA DA MÍDIA
(A)
POPULAR MÍDIA
(B)
FORMATAÇÃO DA MÍDIA
(C)
CRIAÇÃO DA IMAGEM .DD
(D)
APLICAÇÃO DOS BENCHMARK
(E)
AUDITORIA DOS ARQUIVOS RECUPERADOS
(F)

Método

- Ferramentas de data carving:
  - Recurva,
  - Foremost,
  - Scalpel,
  - MagicRescue,
  - PhotoRec,
  - Autopsy.

Método


WIPER/LIMPEZA DA MÍDIA
(A)
POPULAR MÍDIA
(B)
FORMATAÇÃO DA MÍDIA
(C)
CRIAÇÃO DA IMAGEM .DD
(D)
APLICAÇÃO DOS BENCHMARK
(E)
AUDITORIA DOS ARQUIVOS RECUPERADOS
(F)

Métricas auditadas

- Falsos positivos,
- Verdadeiros positivos,
- Tempo de execução.

Resultado: 1° cenário (1.000 arquivos)

Ferramenta	Qtd. de arquivos gerados	Tempo de Execução	Tamanho do diretório recuperado	Falso positivos (%)	Verdadeiro positivos (%)	Verdadeiro positivos repetidos (%)
Recuva	1000	8min 22s	64 MB	$< 1\%$	$> 99\%$	0%
Photorec	768	20min 53s	47,7 MB	$< 1\%$	$> 99\%$	0%
Autopsy	1743	40min 51s	112 MB	$< 1\%$	$> 99\%$	42,4%
Foremost	1563	4min 16s	191,2 MB	38,7%	61,3%	0%
Scalpel	25077	1h 37min 24s	14,5 GB	100%	0%	0%
MagicRescue	162	8min 10s	56,8 MB	94%	6%	0%

Resultado: 2° cenário (16.000 arquivos)

Ferramenta	Qtd. de arquivos gerados	Tempo de Execução	Tamanho do diretório recuperado	Falso positivos (%)	Verdadeiro positivos (%)	Verdadeiro positivos repetidos (%)
Recuva	13952	13h 14min 54s	117 GB	1,1%	98,8%	0,05%
Photorec	14830	06min 50s	5 GB	8,2%	>91,7%	0%
Autopsy	32149	14h 20min 3s	193 GB	10%	90%	44,5%
Foremost	20150	4min 6s	2,2 GB	65,4%	34,5%	0%
Scalpel	508482	13h 54min 15s	100,9 GB	> 99%	< 1%	0%
MagicRescue	15950	2h 4min 47s	2,2 GB	80,1 %	19,9%	0%

Conclusão

- Recuva e PhotoRec - consistentes em relação à quantidade de arquivos recuperados

- Scalpel e MagicRecue – maior quantidade de falsos positivos.

Trabalhos Futuros

- Ampliar escopo (softwares e métricas).
- Ampliar o dataset.
- Propor cenários focados em uma extensão.
- Explorar técnicas de Inteligência Artificial.

Obrigado!

- Rubens K. P. Silva
- rkps@ecomp.poli.br