

Pensamento computacional e IHD: discussões sobre um *workflow* para um estudo de caso

Josiney de Souza^{1,2}, Gabriel S. Hermida², Jefferson L. S. A. Pereira²,
Lucas S. dos Santos², Marcos S. Sunye²

¹IFC *campus* Brusque
Av. Hugo Schlosser, 605 – Jardim Maluche, Brusque – SC, 88354-300

²Departamento de Informática da UFPR
Av. Cel. Francisco H. dos Santos, 100 – Jardim das Américas, Curitiba – PR, 81530-000

josiney.souza@ifc.edu.br, {gash18, jlsap17, lss16, sunye}@inf.ufpr.br

Abstract. *Nowadays, computers are present in people's daily lives and in sciences. This means that computational thinking is important for scientists in their research. With four key techniques (decomposition, pattern recognition, abstraction, algorithms), it promotes logical reasoning and critical thinking in problem solving. Thus, this paper studies and discusses how computational thinking and Human-Data Interaction (HDI) relates to information flows, provenance and interaction of user with data. A case study of a PhD thesis and the construction of a workflow to synthesize data from that research are made.*

Resumo. *Hoje em dia, os computadores estão presentes na rotina das pessoas e nas ciências. Isso significa que o pensamento computacional é importante para os cientistas em suas pesquisas. Com quatro pilares (decomposição, reconhecimento de padrões, abstração, algoritmo), promove o raciocínio lógico e o senso crítico na resolução de problemas. Assim, este trabalho estuda e discute como o pensamento computacional e a Interação Humano-Dados (IHD) se relaciona com fluxo de informações, proveniência dos dados e interação dos usuários com os dados. Foi feito um estudo de caso de uma tese de doutorado e a construção de um workflow para sintetizar dados da referida pesquisa.*

1. Introdução

Hoje em dia, como os computadores fazem parte da rotina das pessoas, o pensamento computacional é importante para lidar com problemas e sistemas complexos. Principalmente nas ciências, onde se usa o método científico para investigações diversas, o pensamento computacional se faz presente.

Essa técnica envolve dividir um problema maior em partes menores e mais simples de se gerenciar (**decomposição**), buscar por similaridades e reaproveitamento de soluções (**reconhecimento de padrões**), criar um modelo simplificado com as informações mais relevantes (**abstração**) e construir um fluxo de passos que resolve o problema (**algoritmo**) [Learning 2018].

Uma das formas de se estruturar o pensamento computacional é através dos fluxos de trabalho ou *workflows*, geralmente apresentados como diagramas no formato

de grafos dirigidos. Neles, cada nodo representa uma decomposição de tarefas; existem caracterizações e perfis de execuções de tarefas (reconhecimento de padrões); sua representação mostra os aspectos mais importantes do mapeamento (abstração); e representa um fluxo de instruções finitas (algoritmo).

Apesar de ser associado à Informática, diferentes contextos e iniciativas fazem uso do pensamento computacional, como a Hora do Código [Code.org 2013] ou maratonas de desenvolvimento (*hackathons*) [SBRC 2023, iCities 2023]. Cientificamente, existem *workflows* para previsão de abalos sísmicos e sequenciamento genômico [Juve et al. 2013].

Nestes, uma característica importante é a proveniência dos dados (*provenance*), que permite rastrear os sucessos e insucessos [Cuevas-Vicenttín et al. 2012, Costa et al. 2013]. Com proveniência de dados, a ciência ganha reprodutibilidade dos experimentos. Outra característica importante é a tomada de decisões de parada, correções e reexecuções pelos cientistas durante as execuções (*Human-in-the-loop*) [Mattoso et al. 2015].

Mesmo com sua importância para a ciência, nem todos os pesquisadores têm o conhecimento necessário para empregar o pensamento computacional. O que se pode fazer com dados adquiridos em pesquisas (construir, adaptar ou usar conjuntos de dados), deve focar em ferramentas e *workflows* além de considerar a organização, formatação, limpeza e correção dos dados para melhor visualização e apresentação [Locke and Heppler 2018].

A Interação Humano-Dados (IHD) é importante nessas situações pois, além da literacia de dados (ler, trabalhar com, analisar, informar e argumentar por meio de dados), se preocupa com as pessoas em todas as fases do processamento dos dados (*Human-in-the-loop*), bem como com os conhecimentos que deles são extraídos e suas visualizações [Brito et al. 2022]. Ou seja, como ajuda a identificar padrões e tendências, o pensamento computacional é inerente à IHD. Ainda, deixa as pesquisas menos sujeitas a erros introduzidos pelos humanos que possam inviabilizá-las.

Há pesquisas que investigam a relação de pensamento computacional e IHD em bases de dados científicos porém nenhuma aplicada à Base de Dados Científicos (BDC) da Universidade Federal do Paraná (UFPR). Trabalhos sobre a BDC descrevem apenas seu planejamento e sua implementação [de Lima et al. 2019, de Lima et al. 2020] mas não a adequação dos *datasets* aos princípios e boas práticas de IHD.

Dado esse contexto, este trabalho fez o estudo de um caso da BDC da UFPR fora da área de saber da Informática para ilustrar como o pensamento computacional e a IHD são importantes e aplicáveis em diferentes situações. Foi explorado conceitos como o fluxo de informações, proveniência dos dados e as interações dos pesquisadores e usuários com os dados para limpeza, formatação e visualização de dados.

As próximas seções trazem o estudo de um caso analisado onde o pensamento computacional e a IHD são importantes (Seção 2); um *workflow* construído sobre o estudo de caso objetivando manutenibilidade e visualização de dados (Seção 3); e os resultados e discussões de seu uso aplicado ao estudo de caso (Seção 4). A Seção 5 conclui este trabalho, seguida pelos agradecimentos (Seção 6) e as referências usadas.

2. Estudo de caso

Um dos conjuntos de dados (*datasets*) existentes na Base de Dados Científicos (BDC) da Universidade Federal do Paraná (UFPR) é “Entre o Amor Venéreo e o Útero: Desdobramentos e Controles de Desgovernadas na Condução da Visita Íntima na Penitenciária Feminina do Paraná” [Gitirana 2021, Gitirana 2020]. Os dados são oriundos de uma pesquisa de doutorado de mesmo nome sobre o perfil, características, atuação profissional, procedimentos das visitas íntimas e direitos sexuais das reclusas sobre o ponto de vista de diretoras e agentes técnicas e penitenciárias, identificadas como burocratas.

Esse caso foi escolhido para estudo pois, durante as buscas na BDC, foi um registros encontrados que possuía ao mesmo tempo um banco de dados disponível, um livro de códigos que mapeasse as escolhas feitas e uma documentação associada (a tese). Adicionalmente, por ser um trabalho externo à área de saber da informática, seria possível comparar as técnicas de uso de dados empregadas com as boas práticas existentes.

O *dataset* possui dois arquivos XLSX (planilhas eletrônicas), que atuam como banco de dados, usados para guardar as respostas recebidas por formulário eletrônico: um para as respostas das diretoras e outro para as das demais burocratas. As respostas quantitativas foram mapeadas para o banco de dados conforme um livro de códigos e as respostas discursivas foram mantidas conforme respondidas.

Porém o *dataset* possui inconsistências evitáveis com IHD e pontos de melhorias que o pensamento computacional permite explorar. Conforme [Locke and Heppler 2018], o planejamento de *datasets* deve ser focado em ferramentas e *workflows*, além de considerar a organização, formatação e correção para a preparação e o uso dos dados.

Como o estudo de caso é feito sobre uma tese fora da área de saber da Informática, assume-se que não foram usados outros bancos de dados além das planilhas eletrônicas, nem técnicas de programação ou de automatizações. Também se assume que o fluxo de trabalho da pesquisa conduzida foi manual (definição das perguntas, aplicação do questionário, tratamento dos dados e análise/síntese), bem como o tratamento, manipulação e interação com os dados coletados.

O estudo de caso focou na análise do Quadro 5 da tese de doutorado, que sintetiza as “respostas das burocratas entrevistadas sobre a autorização da direção da PFP para a realização de visitas íntimas”. O Quadro pode ser visto na captura de tela transformada na Figura 1. No Quadro, há o percentual de respostas “Sim”, “Não” e “Não sei” das burocratas sobre a ciência de visita íntima às mulheres reclusas por pessoas externas ao sistema prisional, por pessoas presas masculinas de outras penitenciárias e por homoafetividade.

Ao se analisar os bancos de dados em formato de planilhas, por exemplo, se constatou que existe uma célula vazia referente à questão da visita íntima homoafetiva no arquivo das agentes penitenciárias e técnicas. A se considerar o restante do conjunto de dados, é possível que esse valor tenha sido apagado mesmo que involuntariamente de forma manual. Ou seja, há uma inconsistência de tipo de dados na base.

Também alguns mapeamentos de respostas são diferentes entre os bancos de dados. No banco de diretoras, mapeia-se 1 para “Não”, 2 para “Sim” e 3 para “Não sabe”. No das agentes, o mapeamento é 1 para “Não sabe”, 2 para “Sim” e 3 para “Não” (primeiro e terceiro valores invertidos). Como os mapeamentos são diferentes, isso requer

Burocracia de rua da PFP	Autorização Tida Direção para visitas íntimas previstas na Portaria 499 de 2014 do DEPEN-PR								
	Visita íntima de mulheres privadas de liberdade			Visita íntima entre pessoas presas*			Visita íntima homoafetiva**		
	Sim	Não	Não sei dizer	Sim	Não	Não sei dizer	Sim	Não	Não sei dizer
Diretoras	100%	0	0	85,7%	14,3%	0	85,7%	14,3%	
Agentes penitenciárias e agentes técnicas	100%	0	0	72%	7%	13%	62,5%	8,7%	17,4%

Figura 1. Captura de tela do Quadro 5 da tese de doutorado (sumarização das respostas das burocratas entrevistadas sobre a autorização da direção da PFP para a realização de visitas íntimas). Fonte: Gitirana (2020)

uma atenção maior na sumarização e análise dos dados ao manipula esse ou aquele banco de dados para se evitar contabilização de uma categoria em outra.

Outro ponto é o campo de “Não sei dizer” estar vazio para a visita íntima homoafetiva referente às diretoras no Quadro 5. Apesar da somatória das demais opções dar 100% dos casos, se assim não o fosse, pode ser que algum dado importante deixaria de ser contabilizado ou exibido; induzindo a erros de conclusões.

Ainda no mesmo Quadro 5, foram usados valores em percentual para indicar os respondentes mas, como a amostra de diretoras é pequena (sete respostas), qualquer mudança para mais ou para menos impacta muito na contagem. Nesse caso, seria interessante ter não só os valores em percentual mas também os valores absolutos.

Mais uma situação é que, nos arquivos XLSX, quando uma questão não era respondida, se atribuía valor 0. Então uma métrica importante de se constar no Quadro 5 seriam as abstenções; tanto que, por exemplo, logo abaixo do Quadro 5 na tese, a autora fez a ressalva de que 4% (duas pessoas) e 8,7% (quatro pessoas) das agentes não responderam sobre as visitas masculinas de outras penitenciárias e homoafetiva, respectivamente.

Por fim, na montagem do questionário on-line, algumas questões de múltipla escolha e resposta única tiveram mais de uma resposta. Tecnicamente, é possível que o formulário tenha sido construído com caixas de seleção (*checkboxes*) em vez de múltipla escolha (*radio buttons*). Isso levou a 4% das agentes (duas pessoas) a marcarem simultaneamente “Sim” e “Não”, conforme ressalvas da autora da tese.

3. Workflow para automatizar síntese de dados

Devido às situações discutidas na seção anterior, foi criado um *workflow* para automatizar a construção do Quadro 5 e demonstrar a viabilidade das boas práticas de IHD e do pensamento computacional. As técnicas utilizadas objetivaram melhorias na visualização de dados e facilidade de manutenção independente do número de respostas existentes.

Conforme [Bharathi et al. 2008], um *workflow* científico possui cinco tipos dife-

rentes de conjunções entre dados e tarefas a se executar: processo, *pipeline*, distribuição de dados, agregação de dados e redistribuição de dados. A Figura 2 mostra as diferentes conjunções.

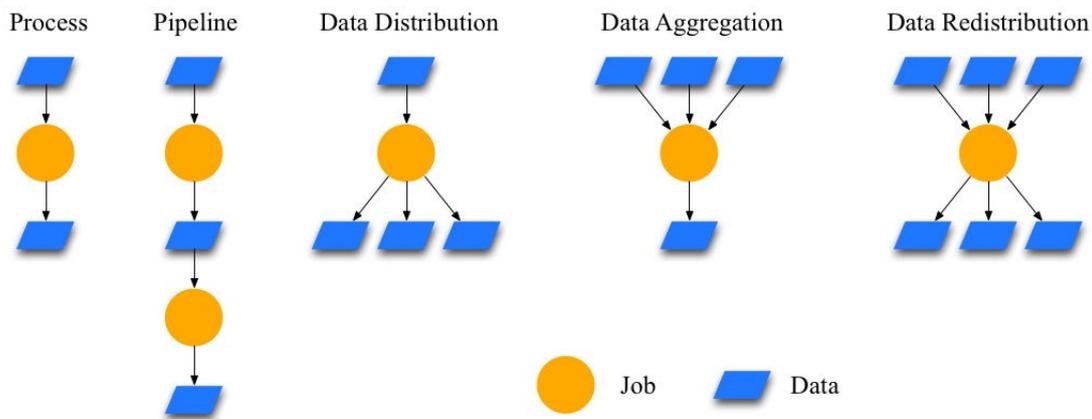


Figura 2. Tipos de estruturas de conexão entre dados e tarefas de processamento. Fonte: Bharathi et al. (2008)

O processo recebe dados, os processa e os transforma. No *pipeline*, a saída de um processo é a entrada do próximo encadeado. Na distribuição de dados, se divide os dados originais para serem usados em outros processamentos em separado. A agregação de dados, condensa os dados separados em uma única saída ao fim do processamento. Na redistribuição de dados, múltiplas entradas são separadas e redivididas em outras saídas.

O *workflow* possui quatro *scripts* para realizar cinco passos diferentes, um para cada nível horizontal da Figura 3. Do primeiro para o segundo nível e do segundo para o terceiro nível, considera-se um processo de distribuição de dados para tratamento em paralelo dos bancos de dados e de cada questão de interesse. Do terceiro para o quarto nível e do quarto para o quinto nível há uma agregação de dados visando a condensação dos resultados de cada passo executado em paralelo.

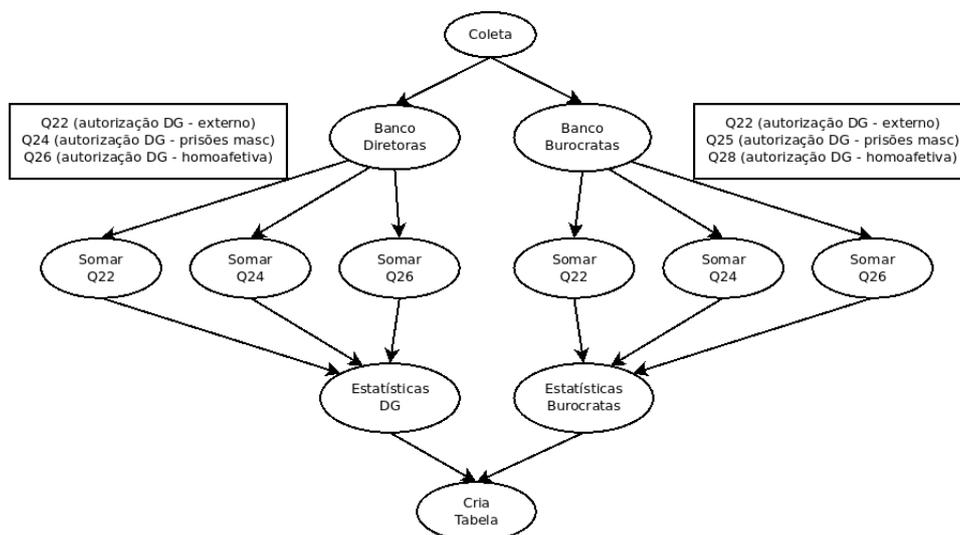


Figura 3. Workflow construído para a síntese automatizada de informações

As questões do formulário on-line aplicado na tese relacionadas ao Quadro 5 respectivamente são as 22, 24 e 26 do banco de dados das diretoras e as questões 22, 25 e 28 do banco de dados das agentes originais. Foi usada programação de *scripts* BASH para a automatização da coleta dos valores de respostas dos bancos de dados e linguagem \LaTeX para se reconstruir o Quadro 5 de forma automatizada. As tarefas executadas em cada nível ou passo são descritas nas seções abaixo.

3.1. Passo 1: Coleta de dados

Neste passo não foi criado nem usado *scripts*, aproveitando-se os bancos de dados em formato XLSX disponibilizados na BDC a partir do *dataset* oriundo da tese de doutorado usada no estudo de caso. Caso não existisse, antes de se executar o *workflow*, seria necessário coletar das respostas a partir de algum meio, como por formulários on-line.

3.2. Passo 2: Transformação dos bancos de dados

Como os bancos de dados originais estão em formato XLSX, foi necessário transformá-los em arquivos de texto puro para se iterar sobre os dados com os *scripts* BASH. A transformação foi realizada com um *script* que utiliza a suíte de escritório LibreOffice a partir da seguinte linha de comando:

```
libreoffice --headless --convert-to csv:"Text - txt - csv  
(StarCalc):35,,0,1,,0,,,,,-1" --outdir $1 $2
```

O comando acima executa o LibreOffice sem interface gráfica e permite converter para o formato de saída CSV considerando os seguintes parâmetros de configuração [LibreOffice 2020]:

1. usar o símbolo “#” como separador de campos/colunas no arquivo de saída (valor “35”, que representa o “#” na tabela ASCII);
2. com a codificação de caracteres do sistema (primeiro “0”);
3. a partir da primeira linha do arquivo original (valor 1);
4. usar o idioma padrão configurado no sistema (segundo “0”); e
5. fazer a conversão para todas as planilhas existentes (valor “-1”).

Demais valores ausentes entre as vírgulas não alteram o comportamento esperado da conversão, sendo configurados com seus valores padrão. Os valores \$1 e \$2 ao fim do comando representam o local de destino do arquivo convertido e a localização do arquivo original, respectivamente, e estão no formato de argumentos passados via linha de comando na chamada de execução do *script*.

3.3. Passo 3: Sumarização das questões

Como cada questão possui mais de um respondente, é necessário contar quantos votos “Não”, “Sim” e “Não sei” foram obtidos. Conforme livro de códigos, no banco de dados de diretoras, as respostas foram mapeadas respectivamente para os valores 1, 2 e 3; no banco de dados de agentes, foram mapeadas respectivamente para 3, 2, 1. O *script* abaixo faz a contagem de cada valor esperado para o banco de dados de diretoras:

```
for VALOR in $QUESTAO; do  
    if [[ $VALOR == 1 ]]; then NAO=$(( $NAO + 1 ))  
    elif [[ $VALOR == 2 ]]; then SIM=$(( $SIM + 1 ))
```

```

elif [[ $VALOR == 3 ]]; then NSABE=$(( $NSABE + 1 ))
elif [[ $VALOR == 0 ]]; then NRESP=$(( $NRESP + 1 ))
elif echo $VALOR | grep -q ";";
then
    echo "Votaram em mais de uma opcao: $VALOR"
else
    echo "Valor \'$VALOR\' invalido"
fi
done

```

A diferença de tratamento para o banco de dados de agentes é apenas o mapeamento de respostas “Não” e “Não sei”. Além desses, o *script* também faz a contagem dos valores 0, quando a questão não foi respondida; a contagem de respostas com “;” (ponto e vírgula), quando uma questão de múltipla escolha de resposta única foi respondida com mais uma opção, por exemplo, “Sim” e “Não” simultaneamente; e a emissão de uma mensagem de aviso caso a resposta dada não esteja em nenhuma das categorias anteriores.

3.4. Passo 4: Gerar estatísticas das contabilizações

Após obter os números totais de votos para cada questão de interesse, faz-se um resumo dos dados contabilizados. O *script* abaixo exibe as estatísticas para cada uma das questões contabilizadas:

```

echo "=== Resultados para $NQUESTAO ($4) ==="
echo "Quantidade de NAO: $NAO"
echo "Quantidade de SIM: $SIM"
echo "Quantidade de NSABE: $NSABE"
echo "Quantidade de NRESP: $NRESP"

```

3.5. Passo 5: Criação de tabela

Previamente, foi criada uma tabela base \LaTeX com símbolos genéricos (*tokens*) que são substituídos no último passo do *workflow* após todos os cálculos anteriores realizados. A Figura 4 ilustra parte da tabela base \LaTeX e dos *tokens* utilizados. Cada *token* está dividido em três partes, separadas por um traço no padrão “PARTE1-PARTE2-PARTE3” (exemplo “DIR-EXT-SIM”).

Autorização Tida Direção para visitas íntimas previstas na Portaria 499 de 2014 do DEPEN-PR								
	Visita íntima de mulheres privadas de liberdade				Visita íntima entre pessoas presas*			
	Sim	Não	Não sei dizer	Não resp.	Sim	Não	Não sei dizer	
Diretoras	DIR-EXT-SIM PDIR-EXT-SIM%	DIR-EXT-NAO PDIR-EXT-NAO%	DIR-EXT-NSABE PDIR-EXT-NSABE%	DIR-EXT-NRESP PDIR-EXT-NRESP%	DIR-MASC-SIM PDIR-MASC-SIM%	DIR-MASC-NAO PDIR-MASC-NAO%	DIR-MASC-NSABE PDIR-MASC-NSABE%	I PD
Agentes penitenciárias e agentes técnicas	BUR-EXT-SIM PBUR-EXT-SIM%	BUR-EXT-NAO PBUR-EXT-NAO%	BUR-EXT-NSABE PBUR-EXT-NSABE%	BUR-EXT-NRESP PBUR-EXT-NRESP%	BUR-MASC-SIM PBUR-MASC-SIM%	BUR-MASC-NAO PBUR-MASC-NAO%	BUR-MASC-NSABE PBUR-MASC-NSABE%	B PB

Figura 4. Recorte representativo da tabela base \LaTeX usando *tokens*

Na primeira parte, identifica-se tanto a categoria de entrevistadas (“DIR” para diretoras e “BUR” para demais burocratas) quanto o tipo de sumarização, se valor absoluto de contagem de respostas (“DIR”) ou se valor percentual (“PDIR”). Na segunda parte, identifica-se o tipo de visitas; se por pessoas externas ao sistema prisional (“EXT”), por homens/pessoas masculinas de outras penitenciárias (“MASC”) ou por visitas homoafetivas (“HOMO”). Na terceira parte, identifica-se a resposta dada: “Sim”, “Não”, “Não sei” e “Não respondeu” (“SIM”, “NAO”, “NSABE” e “NRESP” respectivamente).

4. Resultados e discussões

Abaixo, na Tabela 1, apresenta-se o resultado da execução do *workflow* construído considerando-se as bases de dados originais da pesquisadora. A tabela gerada está em formato $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ pronta para ser adicionada a documentos. É possível verificar que todos os campos presentes no Quadro 5 foram reproduzidos e que o novo campo “Não resp.”, para a visualização das abstenções, foi introduzido.

Tabela 1. Reprodução do Quadro 5 via *workflow*

Burocracia de rua da PFP	Autorização da Direção para visitas íntimas previstas na Portaria 499 de 2014 do DEPEN-PR											
	Visita íntima de mulheres privadas de liberdade				Visita íntima entre pessoas presas*				Visita íntima homoafetiva**			
	Sim	Não	Não sei dizer	Não resp.	Sim	Não	Não sei dizer	Não resp.	Sim	Não	Não sei dizer	Não resp.
Diretoras	7 100.00%	0 0%	0 0%	0 0%	6 85.71%	1 14.28%	0 0%	0 0%	5 71.42%	2 28.57%	0 0%	0 0%
Agentes penitenciárias e agentes técnicas	46 100.00%	0 0%	0 0%	0 0%	33 71.73%	3 6.52%	6 13.04%	2 4.34%	30 65.21%	4 8.69%	7 15.21%	4 8.69%

Com o *workflow*, há ganho de escala. Indêpende da quantidade de respostas, os passos executados serão os mesmos e feitos em lote ainda que houvessem mais respostas ao questionário on-line para qualquer categoria de respondentes. Também se poderia fazer correções e ajustes em lote nas respostas sem trabalho adicional para a pesquisadora.

Outra vantagem é a redução das chances de introdução de erros por via humana. Se comparar as respostas “Sim” e “Não” do Quadro 5 com o banco de dados e a reprodução com *workflow* para a visita íntima homoafetiva sob a visão das diretoras, nota-se que o original possui erro de contabilização; possivelmente um copia-e-cola da categoria de visita íntima entre pessoas presas.

Mais divergências estão em “Sim” e “Não sei dizer” para as agentes sobre as visitas íntimas homoafetivas. No caso do “Sim”, o original contabiliza menos que 30 respondentes; porém 29 e 28 respondentes representariam a 63,04% e 60,87% respectivamente e não os 62,5% informados. No caso do “Não sei dizer”, o original contabiliza um respondente a mais (oito, totalizando 17,39%).

Uma desvantagem do uso do *workflow* como foi construído é o conhecimento prévio de algumas técnicas tipicamente da área de Informática para a manutenção ou

expansão do *workflow*: programação de computadores, para entender os códigos dos *scripts*); sistemas operacionais (para executar os *scripts*); e \LaTeX , para adequar a tabela base do *workflow*. Ou seja, apesar do ganho de tempo para a pesquisa, a curva de aprendizado pode ser grande.

5. Conclusão

No mundo atual, onde a computação se faz presente no dia a dia, o pensamento computacional é uma ferramenta para abordar problemas das mais variadas áreas. A partir da decomposição, do reconhecimento de padrões, da abstração e de algoritmos, é possível modelar situações simples ou complexas que necessitem raciocínio lógico e senso crítico.

Este trabalho estudou a importância do pensamento computacional e de boas práticas de IHD para as ciências nos dias atuais, discutindo o fluxo de informações e a proveniência de dados. Analisou erros cometidos pela falta de uso dessas técnicas a partir de um quadro sumarizador de uma tese de doutorado. Também construiu um *workflow* e o aplicou ao conjunto de dados da tese para demonstrar os benefícios do tratamento e melhoria da visualização de dados independentemente de área do conhecimento.

Por outro lado, o *workflow* construído tratou de apenas um quadro sumarizador da tese de doutorado usada como base; assim, trabalhos futuros podem envolver expansão do *workflow* para reproduzir outros quadros sumarizadores da tese. Ainda, foi analisado apenas um trabalho da BDC; então outro caminho futuro é realizar a análise de outros *datasets* da BDC ou de outros repositórios científicos sob os mesmos pontos de vista.

6. Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e Programa de Excelência Acadêmica (PROEX). Agradece-se ao DInf da UFPR pela vivência em seu ambiente acadêmico e bolsa de doutorado. Agradece-se também ao IFC e ao IFC *campus* Brusque por oportunizar o afastamento do servidor para capacitação.

Referências

- Bharathi, S., Chervenak, A., Deelman, E., Mehta, G., Su, M.-H., and Vahi, K. (2008). Characterizing of scientific workflows. In *Third Workshop on Workflows in Support of Large-Scale Science*, pages 1–10. IEEE.
- Brito, L., França, J., Morais, L., Queiroz, C., Dias, A., and Vivacqua, A. (2022). Workshop investigações em interação humano-dados – wide. In *Anais Estendidos do XXI Simpósio Brasileiro de Fatores Humanos em Sistemas Computacionais*, pages 5–8, Porto Alegre, RS, Brasil. SBC.
- Code.org (2013). Hour of code. Disponível em: <https://hourofcode.com/br>. Acesso em: 20 de jun. de 2023.
- Costa, F., Silva, V., de Oliveira, D., Ocaña, K., Ogasawara, E., Dias, J., and Mattoso, M. (2013). Capturing and querying workflow runtime provenance with prov: a practical approach. In *EDBT '13: Proceedings of the Joint EDBT/ICDT 2013 Workshops*, pages 282—289. ACM.

- Cuevas-Vicentín, V., Dey, S., Köhler, S., Riddle, S., and Ludäscher, B. (2012). Scientific workflows and provenance: Introduction and research opportunities. In *Datenbank Spektrum*, pages 193—203. Springer.
- de Lima, K. C. R., de Araújo, P. C., and Estevão, J. S. B. (2019). A implantação da base de dados científicos (bdc) da universidade federal do paran  (ufpr). In *IX Confer ncia Internacional sobre Bibliotecas e Reposit rios Digitais da Am rica Latina BIREDIAL – ISTECA Anais das sess es tem ticas*, pages 195–202. Servicio de Difusi n de la Creaci n Intelectual es el Repositorio Institucional de la Universidad Nacional de La Plata. Dispon vel em: <http://sedici.unlp.edu.ar/handle/10915/86755>. Acesso em: 13 de nov. de 2023.
- de Lima, K. C. R., Estev o, J. B., Ara jo, P. C. d., and Gonalves, L. H. (2020). A base de dados cient ficos da universidade federal do paran : do planejamento   implantao. *Ci ncia da Informao*, 48(3).
- Gitirana, J. (2021). Entre o amor ven ris e o  tero: Desdobramentos e controles de des governadas na conduo da visita  tima na penitenci ria feminina do paran . Dispon vel em: <http://dx.doi.org/10.5380/bdc/63>. Acesso em: 19 de jun. de 2023.
- Gitirana, J. H. S. (2020). *Entre o amor ven ris e o  tero : desdobramentos e controle de des governadas na conduo da visita  tima na penitenci ria feminina do Paran *. PhD thesis, Universidade Federal do Paran . Dispon vel em: <https://hdl.handle.net/1884/70370>. Acesso em: 19 de jun. de 2023.
- iCities (2023). Smart city expo curitiba 2023 ter  como tema conectividade global e cidades sem fronteiras. Dispon vel em: <https://www.icities.com.br/2022/07/30/smart-city-expo/>. Acesso em: 20 de jun. de 2023.
- Juve, G., Chervenak, A., Deelman, E., Bharathi, S., Mehta, G., and Vahi, K. (2013). Characterizing and profiling scientific workflows. In *Future Generation Computer Systems*, pages 682–692. Elsevier.
- Learning, B. (2018). What is computational thinking? Dispon vel em: <https://www.bbc.co.uk/bitesize/topics/z7tp34j>. Acesso em: 28 de jun. de 2023.
- LibreOffice (2020). Csv filter parameters. Dispon vel em: https://help.libreoffice.org/latest/en-GB/text/shared/guide/csv_params.html. Acesso em: 13 de nov. de 2023.
- Locke, B. T. and Heppler, J. A. (2018). Teaching data literacy for civic engagement: Resources for data capture and organization. In *KULA: knowledge creation, dissemination, and preservation studies*, pages 1–5. University of Victoria Libraries.
- Mattoso, M., Dias, J., Oca na, K. A., Ogasawara, E., Costa, F., Horta, F., Silva, V., and de Oliveira, D. (2015). Dynamic steering of hpc scientific workflows: A survey. In *Future Generation Computer Systems*, pages 100–113. Elsevier.
- SBRC (2023). Hackathon smartness / 5g dataset challenge. Dispon vel em: <https://sbrc.sbc.org.br/2023/hackathon/>. Acesso em: 20 de jun. de 2023.