

BASES TEÓRICAS PARA A APLICAÇÃO DE MINERAÇÃO DE DADOS NA PESQUISA EM DESIGN

RAFAEL PEDUZZI GOMES¹; YGOR CORRÊA²; VINICIUS GADIS RIBEIRO³

¹Universidade Federal de Pelotas – rafaelpeduzzi@gmail.com

²UniRitter Laureate International Universities – correaygorprof@gmail.com

³Universidade Federal do Rio Grande do Sul – vinicius.gadis@ufrgs.br

1. INTRODUÇÃO

O presente trabalho tem como ponto de partida o interesse na Pesquisa em Design como objeto de estudo, que se configura como um fato recente considerado como “sinal de boa saúde” no campo (LLOYD, 2017). Ou seja, pretende-se ampliar o conhecimento a respeito da área, aprofundando o olhar sobre temáticas, métodos e técnicas abordadas em produções científicas atuais.

Tem-se como premissa que a Pesquisa em Design está em expansão (GEMSER; DE BONT, 2016), com número crescente de periódicos e congressos, e com seus dados cada vez mais disponíveis em formato digital. Neste sentido, são positivos os esforços para, além de organizar e mapear o conhecimento que é produzido, buscar novas técnicas para extrair conteúdos úteis da grande quantidade de informação na área. Percebe-se como possibilidade pouco explorada a aplicação de mineração de dados para descobrir conhecimento em grandes bases de dados de pesquisa. Aponta-se inicialmente que os resultados de tal abordagem podem colaborar para encontrar possíveis tendências quanto a métodos, rastrear tópicos recorrentes ou relacionados, além de instituições ou localizações mais focadas em técnicas específicas de pesquisa.

Este trabalho busca investigar a aplicabilidade e o potencial das técnicas de mineração de dados para descobrir conhecimento sobre a Pesquisa em Design. São investigadas, a partir de pesquisa bibliográfica, técnicas de mineração de textos, análise de agrupamento e análise de associação, a fim de delinear as mais adequadas para um futuro estudo aplicado. Assim, a questão de pesquisa pode ser sintetizada da seguinte forma: como a mineração de dados pode ser aplicada para descoberta de conhecimento sobre a Pesquisa em Design e quais técnicas seriam mais adequadas para isso? Pretende-se para estudos futuros, a partir dessas bases teóricas, gerar visualizações de dados e interpretações que possam representar a Pesquisa em Design na atualidade.

2. METODOLOGIA

Este estudo pode ser caracterizado como uma pesquisa qualitativa de alcance exploratório. O enfoque qualitativo, a partir de SAMPIERI, COLLADO e LUCIO (2013), envolve um pesquisador que investiga o mundo social e “desenvolve uma teoria coerente com os dados, de acordo com aquilo que observa” (SAMPIERI; COLLADO; LUCIO, 2013, p. 33). Assim, as pesquisas qualitativas têm um processo indutivo de explorar e descrever, indo do particular ao geral. Ainda segundo os autores, é possível no enfoque qualitativo construir perguntas e hipóteses antes, durante e depois da coleta e análise de dados.

Os estudos exploratórios são citados por SAMPIERI, COLLADO e LUCIO (2013, p. 100) como “realizados quando o objetivo é examinar um tema ou problema de pesquisa pouco estudado, sobre o qual temos muitas dúvidas ou que

não foi abordado antes”. Nesse sentido, mesmo que as técnicas de mineração de dados não sejam uma novidade, não foi encontrado um número significativo de aplicações dessas técnicas para buscar conhecimentos em bases de dados de Pesquisa em Design; o trabalho encontrado nesse sentido é citado aqui (RIBEIRO et al., 2013). Dessa forma, a revisão de literatura mostrou apenas orientações e ideias pouco relacionadas ao objeto de estudo. Busca-se, então, intenções que se enquadram na definição de “obter informação sobre a possibilidade de realizar uma pesquisa mais completa relacionada com um contexto particular, [...] identificar conceitos ou variáveis promissoras, estabelecer prioridades para pesquisas futuras” (SAMPIERI; COLLADO; LUCIO, 2013, p. 101).

Dessa forma, busca-se aqui um embasamento teórico que suporte a possível aproximação entre técnicas de mineração de dados, já utilizadas previamente no âmbito da pesquisa científica, com a Pesquisa em Design, apontada como nova e com divergências e confusões teóricas (LOVE, 2000; BAYAZIT, 2004; BECCARI, 2012; LLOYD, 2017), ensejando, portanto, uma investigação aprofundada de sua produção teórica e científica.

3. RESULTADOS E DISCUSSÃO

Apresenta-se aqui os resultados iniciais da pesquisa bibliográfica conduzida a fim de investigar as potencialidades da mineração de dados. Num contexto de dados coletados e acumulados de modo cada vez mais acelerado, FAYYAD et al. (1996) postulam a necessidade urgente de novas teorias computacionais e ferramentas para auxiliar humanos a extrair informações úteis do volume crescente de dados digitais. Tais teorias e ferramentas são pertencentes ao campo de *Knowledge Discovery in Databases* (KDD, Descoberta de Conhecimento em Bases de Dados). A mineração de dados é uma parte integral da KDD, definida como o processo geral de conversão de dados brutos em informações úteis (TAN; STEINBACH; KUMAR, 2006, p. 4).

Quanto às etapas desse processo, segundo TAN, STEINBACH e KUMAR (2006), o propósito do pré-processamento é transformar os dados brutos de entrada em um formato adequado para análises, considerando principalmente o tipo de arquivo necessário para executar determinadas operações nas ferramentas computacionais. Os dados podem ser armazenados em diversos formatos, como arquivos simples, planilhas ou tabelas relacionais, e estar em um repositório central ou ainda em múltiplos locais. O pós-processamento, ainda a partir de TAN, STEINBACH e KUMAR (2006), é uma etapa de validação, que pode ter medições estatísticas e testes de hipótese, com a intenção de eliminar resultados não legítimos da etapa de mineração. Um exemplo de pós-processamento proposto é a visualização, que consiste basicamente na conversão das informações em um formato visual, como gráfico ou tabela.

A mineração de dados pressupõe o uso de métodos específicos, que vão além de uma busca por dados, na tarefa de garimpar, explorar, extrair alguma substância de uma determinada fonte. Assim, é definida como o processo de descoberta automática ou semiautomática de informações úteis em grandes bases de dados (TAN; STEINBACH; KUMAR, 2006, p. 2). Suas técnicas buscam padrões e conhecimentos que poderiam ser ignorados em uma busca comum. Os autores comentam que nem todas as tarefas de descoberta de informação são consideradas mineração de dados, por exemplo, buscas na Internet ou em bases de dados são tarefas relacionadas à área da recuperação de dados (TAN; STEINBACH; KUMAR, 2006, p. 3).

ROWLEY (2007) define a relação entre dados, informação, conhecimento e sabedoria: os dados são itens elementares, coletados e armazenados por recursos de Tecnologia da Informação; as informações representam os dados processados com significados e contextos bem definidos; o conhecimento é o padrão ou padrões cuja formulação envolve e relaciona dados e informações já obtidas; por fim, a sabedoria é o conhecimento de especialista e julgamento das questões quanto ao significado e condução da vida. Segundo FAYYAD et al. (1996), um padrão deve ser entendido como um conhecimento representado segundo as normas sintáticas de uma linguagem formal, isto é, reconhecível pelo ser humano. Disso se conclui que, para a descoberta de conhecimento, é necessária a atuação do ser humano nesse processo, observando e interpretando os fatos durante a mineração de dados.

TAN, STEINBACH e KUMAR (2005) propõem a seguinte divisão das tarefas de mineração de dados: (1) Tarefas de Previsão, com objetivo de prever o valor de um atributo a partir dos valores de outros atributos; e (2) Tarefas Descritivas, cujo objetivo é encontrar padrões que resumam os relacionamentos subjacentes nos dados. As tarefas descritivas têm, em geral, natureza exploratória, e com frequência requerem técnicas de pós-processamento para validar e explicar resultados. Conforme FAYYAD et al. (1996), para descoberta de conhecimento, pode-se usar diversos métodos: Classificação, Análise de Agrupamento, Sumarização, Regras de Associação e Análise de Séries Temporais.

Já a mineração de textos é uma tecnologia para descobrir de forma semiautomática padrões, tendências e conhecimentos antes desconhecidos em bases de texto não-estruturado, segundo MALI e ATIQUÉ (2014). Logo, é diferente da mineração de dados, que se aplica principalmente em dados já estruturados. A maioria dos dados a serem abordadas na mineração de texto não estão estruturados, como no caso de textos completos ou resumos de artigos. Então, é necessária uma ferramenta que possa dar suporte à conversão dessa informação em conhecimento, sendo algumas funções da mineração de texto: “busca, classificação, agrupamento, sumarização, monitoramento de informação e extração de informação” (MALI; ATIQUÉ, 2014, p. 209).

Por fim, LAHLOU (2001) coloca que o software não é o método, e que os dados de saída fornecidos pelo software não são a análise. Assim, enfatiza a relevância da interpretação no processo, sendo um processo abduutivo que depende do conhecimento externo do analista sobre a linguagem, o tópico e o software. São, portanto, fatores limitadores, podendo-se concluir que o software é um instrumento para exploração e auxílio na interpretação.

4. CONCLUSÕES

De forma geral, verificou-se que há potencial na mineração de textos como ferramenta para descobrir conhecimentos em bases de dados de pesquisa científica, mesmo que apenas a partir deste primeiro estudo, em estágio inicial de pesquisa. É importante ressaltar que os resultados podem ser mais precisos na medida em que se triangula métodos, como já se referia LAHLOU (2001). Algumas técnicas a serem aplicadas podem ser a análise de agrupamentos, buscando-se detectar estudos que possuem mais afinidade entre si, e análise de associação, a fim de perceber regras que associem termos utilizados em estudos, obtidos a partir da mineração de textos.

A partir do estudo, foram definidos artigos científicos como foco de estudos futuros. Parte-se da premissa de que são um formato priorizado no atual modelo de produção científica segundo RIBEIRO et al. (2013), o qual considera que na

comunicação científica, fase de uma pesquisa em que se divulga os resultados, os trabalhos publicados em periódicos ou eventos são, em geral, os considerados de boa qualidade a partir de avaliação feita pelos seus pares, os quais formam um corpo editorial. Portanto, os artigos se apresentam como produção já referendada por editores de periódicos, organizadores de eventos e pesquisadores da área.

Reitera-se o potencial e a relevância da mineração de dados para descobrir conhecimentos sobre a Pesquisa em Design. Sendo um foco de estudo ainda pouco explorado, espera-se encontrar novas e úteis informações na medida em que se prossegue o estudo neste sentido, contribuindo para um maior conhecimento da área do Design sobre sua própria pesquisa.

5. REFERÊNCIAS BIBLIOGRÁFICAS

BAYAZIT, N. Investigating Design: A Review of Forty Years of Design Research. **Design Issues**, v. 20, n. 1, p. 16–29, 2004.

BECCARI, M. **Articulação simbólica: uma abordagem junguiana aplicada à filosofia do design**. 2012. 380 f. Dissertação (Mestrado em Design) – Setor de Ciências Humanas, Letras e Artes, Universidade Federal do Paraná, Curitiba, 2012.

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. 1996. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 2, Portland, 1996. **Proceedings**... Portland, Oregon, 1:33-47.

GEMSER, G.; DE BONT, C.. Design-Related and Design-Focused Research: A Study of Publication Patterns in Design Journals. **She Ji: The Journal of Design, Economics, and Innovation**. Shanghai, Vol. 2, n. 1, p. 46–58, Maio 2016.

LAHLOU, S. Text mining methods: an answer to Chartier and Meunier. **Papers on Social Representations**, v. 20, n. 38, 2001, p. 1-7.

LOVE, T. Philosophy of design: a meta- theoretical structure for design theory. **Design Studies**, v. 21, n. 3, p. 293–313, 2000.

MALI, M.; ATIQUE, M. Applications of Text Classification using Text Mining. **International Journal of Engineering Trends and Technology**. Índia, v. 13, n. 5. 2014, p. 209-212.

RIBEIRO, V.; SILVEIRA, S.; SILVEIRA, A.; ATKINSON, R.; ZABADAL, J. The use of data mining techniques for defining strategies in scientific communication processes in design journals. **Strategic Design Research Journal**, v. 6, n. 2, 2013, p. 85-94.

ROWLEY, J. The wisdom hierarchy: representations of the DIKW hierarchy. **Journal of Information Science**, v. 33, n. 2, p. 163–180, 2007.

SAMPIERI, R; COLLADO, C; LUCIO, M. **Metodologia de Pesquisa**. Porto Alegre: Penso Editora, 2013.

TAN, P.; STEINBACH, M.; KUMAR, V. **Introduction to Data Mining**. Edinburgo: Pearson Addison-Wesley, 2006.