



Em Questão
ISSN: 1807-8893
ISSN: 1808-5245
emquestao@ufrgs.br
Universidade Federal do Rio Grande do Sul
Brasil

Aplicação de revisão sistemática com suporte de mineração de dados e de textos: o caso do periódico Design Studies

Gomes, Rafael Peduzzi; Ribeiro, Vinicius Gadis; Corrêa, Ygor; Zabadal, Jorge Rodolfo Silva

Aplicação de revisão sistemática com suporte de mineração de dados e de textos: o caso do periódico Design Studies

Em Questão, vol. 25, núm. 3, 2019

Universidade Federal do Rio Grande do Sul, Brasil

Disponível em: <https://www.redalyc.org/articulo.oa?id=465660194008>

DOI: <https://doi.org/10.19132/1808-5245253.156-183>



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial 3.0 Internacional.

Aplicação de revisão sistemática com suporte de mineração de dados e de textos: o caso do periódico Design Studies

Application of systematic review with support of data and text mining: the case of the journal Design Studies

Rafael Peduzzi Gomes 1
rafaelpeduzzi@gmail.com

DOI: <https://doi.org/10.19132/1808-5245253.156-183>

Redalyc: <https://www.redalyc.org/articulo.oa?id=465660194008>

Vinicius Gadis Ribeiro 2
vinicius.gadis@ufrgs.br

Ygor Corrêa 3
correaygorprof@gmail.com

Jorge Rodolfo Silva Zabadal 4
jorge.zabadal@ufrgs.br

Recepción: 19 Septiembre 2018

Aprobación: 20 Febrero 2019

RESUMO:

O emprego de revisão sistemática na academia tem apresentado crescimento considerável. Frequentemente, esta tem sido usada como uma forma de filtragem por meio de palavras-chave. Contudo, é levantada, nesse processo, uma grande quantidade de dados, a partir da qual se acredita que possam ser identificadas relações mais aprofundadas, empregando técnicas de mineração de dados ou textos. Este estudo tem como objetivo identificar e verificar relações entre tópicos de pesquisa, considerando como uma prova de conceito os artigos do periódico *Design Studies*, por meio da aplicação de tecnologias de Descoberta de Conhecimentos em Bases de Dados. Em uma abordagem quantitativa de alcance descritivo, apresenta-se, inicialmente, uma caracterização da amostra de 483 artigos desse periódico a partir de estatística descritiva, abarcando as palavras-chave provenientes dos seus metadados. Em seguida, é usada a ferramenta de mineração de texto *AlchemyAPI*, para extração automática de conceitos e palavras-chave, e a ferramenta *RapidMiner Studio*, para análise de associação a partir dos metadados e dos resultados da mineração de texto. Constatou-se a recorrência de tópicos como o Processo, a Cognição, a Criatividade e a Educação em Design. Foi possível verificar, por meio da ferramenta *AlchemyAPI*, conceitos que não foram expressos nas palavras-chave dos artigos, o que pode ser considerado como uma potencialidade para estudos de revisão sistemática. Assim, espera-se que as técnicas aqui abordadas possam ser mais exploradas em pesquisas de diversas áreas, enquanto suporte válido para revisões sistemáticas de publicações científicas.

PALAVRAS-CHAVE: Revisão sistemática, Mineração de Dados, Mineração de Textos.

ABSTRACT:

The use of systematic review in academia has shown considerable growth. It has often been used as a form of keyword filtering. However, a large amount of data is collected in this process, from which it is believed that more in-depth relationships can be identified, using data or text mining techniques. This study aims to identify and verify relationships between research topics, considering, as a proof of concept, papers in the journal *Design Studies*, through the application of Knowledge Discovery in

NOTAS DE AUTOR

- 1 Mestre; Universidade Federal de Pelotas, Pelotas, RS, Brasil
rafaelpeduzzi@gmail.com
- 2 Doutor; Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil
vinicius.gadis@ufrgs.br
- 3 Doutor; Universidade de Caxias do Sul, Caxias do Sul, RS, Brasil
correaygorprof@gmail.com
- 4 Doutor; Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

Databases' techniques. In a quantitative approach with a descriptive scope, a characterization of the sample of 483 articles is presented initially with descriptive statistics, encompassing keywords from metadata. Then, AlchemyAPI is used for automatic extraction of concepts and keywords, and RapidMiner Studio is used for association analysis from metadata and text mining results. It was found a recurrence of themes in Design such as Process, Cognition, Creativity, and Education. It was possible to verify, through the AlchemyAPI tool, concepts that were not expressed in papers' keywords, which can be considered as a potential for systematic review studies. Thus, it is expected that the techniques applied here may be more explored in research on Design and other areas, as a valid support for systematic reviews of scientific publications.

KEYWORDS: Text mining, Research topics, Design research.

1 INTRODUÇÃO

Nos atuais moldes de produção científica, a fase final, de comunicação dos resultados, acontece prioritariamente na forma de artigos publicados em periódicos ou anais de eventos. Nesse cenário, há o que tem sido denominado como o problema da grande, crescente e incontrolável quantidade de informações de pesquisa, principalmente nos últimos 20 anos, diante do surgimento de novos periódicos e da publicação de milhares de artigos (HIGGINS; GREEN, 2008; PETTICREW; ROBERTS, 2006), sendo 80% da informação em formato textual (SOMASUNDARAM; SHRIVASTAVA, 2011).

Em busca de lidar com o desafio de gerar conhecimento a partir do excesso de informação em diversos âmbitos (SÉRGIO; SILVA; GONÇALVES, 2016), surgem as técnicas de Descoberta de Conhecimento em Bases de Dados (do inglês, *Knowledge Discovery in Databases* - KDD), englobando a mineração de dados e a mineração de textos (TAN; STEINBACH; KUMAR, 2006). O processo de KDD busca desvendar padrões e tendências no extenso volume de dados gerados na sociedade contemporânea. Ainda nesse contexto, se considera que a KDD pode servir como um suporte a revisões sistemáticas em estudos que se debruçam sobre a produção científica, mais especificamente sobre tópicos de pesquisa, ou seja, termos que representam um dos assuntos associados a um determinado documento (BORGES; NOGUEIRA; BARBOSA, 2015). Dessa forma, pode-se, a partir da necessidade de avaliação de uma ciência institucionalizada (ALVAREZ; CAREGNATO, 2017), ampliar a compreensão de uma área de conhecimento científico, abordando os principais tópicos, as relações entre tópicos e as tendências das pesquisas, e contribuir com pesquisadores quanto à seleção de publicações de seu interesse e à fundamentação para novas pesquisas, progredindo teórica e metodologicamente.

Neste estudo, os esforços são direcionados para o campo do Design, no qual se entende, a partir de Love (2000), que há uma imprecisão quanto à aderência de bases teóricas e métodos, além de uma multiplicidade de teorias que dialogam com o Design e diversas perspectivas na atualidade (BAYAZIT, 2004; BUCHANAN, 2001; CROSS, 2001; LLOYD, 2017), no que já foi chamado de “confusão teórica” (BECCARI, 2012). A pesquisa acadêmica e científica em Design encontra-se em expansão (GEMSER; DE BONT, 2016) e mais relevante do que nunca (LLOYD, 2017), sendo crescente o número de periódicos e anais de congressos com produção disponível em formato digital, o que possibilita estudos a partir desses dados.

Além disso, a Pesquisa em Design como objeto de estudo é um fato recente, considerado como “sinal de boa saúde” (LLOYD, 2017). Bayazit (2004) aborda a necessidade de pesquisa extensa sobre a Ciência, as Metodologias e a Pesquisa em Design, enquanto Cross (2007) aponta uma escala internacional da Pesquisa em Design, com novos periódicos e conferências, bem como um crescente interesse em pós-graduação na área, também indicado por Lloyd (2017). Assim, tendo em conta que a produção de conhecimento acadêmico-científico no Design é interdisciplinar, recente e não apresenta uma tradição de metodologias próprias (FRIEDMAN, 2001; GEMSER et al., 2012; GEMSER; DE BONT, 2016; RIBEIRO; SILVEIRA; SILVEIRA; ATKINSON; ZABADAL, 2013; RIBEIRO; SILVEIRA; MANINI; BARROSO; ZABADAL, 2016), tendo uma história longa, mas não robusta (OWEN, 1998), com cerca de 50 anos (LLOYD, 2017), é relevante verificar como essas questões teóricas apresentam-se nos tópicos de pesquisa mais recentes na área.

Assim, foi delineada uma proposição metodológica de alcance descritivo e de enfoque quantitativo (SAMPIERI; COLLADO; LUCIO, 2013), que consiste em uma revisão sistemática de literatura ancorada na extração de conceitos, técnica relacionada à mineração de textos, e na análise associativa, relacionada à mineração de dados. Foram escolhidos, como foco de pesquisa, artigos científicos na área do Design, considerando seu caráter de produção já referendada e priorizada na ciência (RIBEIRO; SILVEIRA; SILVEIRA; ATKINSON; ZABADAL, 2013). Escolheu-se como objeto de estudo o periódico *Design Studies*, apontado como um dos melhores (GEMSER *et al.*, 2012), mais interdisciplinares (CROSS, 2007), citados (CHAI; XIAO, 2012) e bem ranqueados em Design. O periódico está disponível para acesso digital, viabilizando a investigação. Busca-se ampliar o escopo de pesquisa (2001-2016) do *Design Studies*, que mostrou potencial em estudos prévios (CHAI; XIAO, 2012; RIBEIRO *et al.*, 2013; RIBEIRO *et al.*, 2016), e continuar a exploração quantitativa da Pesquisa em Design, apontada como quase inexistente e como útil em outros campos (CHAI; XIAO, 2012).

Dessa forma, objetiva-se realizar uma investigação usando as tecnologias de mineração de dados e de textos como suporte a uma revisão sistemática, na qual são levantadas as palavras-chave dos metadados dos artigos, sendo essas comparadas às palavras-chave e aos conceitos retornados pela ferramenta de mineração de texto *AlchemyAPI*. Por fim, por meio de uma análise associativa, com suporte do software *RapidMiner Studio*, são analisadas as principais regras de associação e coocorrências entre tópicos. Considera-se que essa análise pode encontrar resultados que contribuam tanto com uma progressão teórica na Pesquisa em Design quanto com novos métodos e técnicas no âmbito da Ciência da Informação, testando e avaliando sua aplicação e as ferramentas utilizadas, o que pode ser estendido a estudos similares em outras áreas científicas.

A partir disso, o artigo está estruturado da seguinte forma: na seção 2, são explanados os principais pressupostos teóricos deste estudo, enquanto na seção 3 são detalhados os procedimentos metodológicos. Na seção 4, são apresentados os resultados da caracterização da amostra e das aplicações de KDD e, na seção final, são explicitadas as considerações finais do estudo. A seguir, são abordados os conceitos de revisão sistemática, mineração de dados e de textos.

2 REVISÃO SISTEMÁTICA DE LITERATURA E MINERAÇÃO DE TEXTOS

As revisões de literatura podem ter diversos propósitos, como: examinar velhas teorias e propor novas; resumir e considerar se há evidência a respeito de um tópico; guiar pesquisadores no planejamento de estudos futuros; examinar de forma metódica os conflitos de diferentes estudos sobre um mesmo problema (PETTICREW; ROBERTS, 2006). Por exemplo, Akobeng (2005) coloca que profissionais da saúde sempre usaram artigos de revisão como resumos de evidências em certos tópicos, o que pode ser expandido para as Ciências Sociais a partir de Petticrew e Roberts (2006), que colocam essa como uma abordagem científica tradicional. Assim, define-se que revisões sistemáticas são um tipo particular de revisão de literatura, que coleta e analisa de forma crítica múltiplos estudos a partir de uma metodologia estruturada (AKOBENG, 2005) e claramente explicitada (HIGGINS; GREEN, 2008), a fim de estabelecer uma síntese que mostre quais descobertas são relevantes e o que pode ser utilizado ou descartado (PETTICREW; ROBERTS, 2006), bem como as questões que podem ser respondidas a partir dos estudos abordados.

Dessa forma, definem-se previamente objetivos e critérios, em uma metodologia explícita, formal e reproduzível, a fim de minimizar a possibilidade de viés (PETTICREW; ROBERTS, 2006; HIGGINS; GREEN, 2008), ou seja, a possibilidade de interferências indesejadas das escolhas do pesquisador nos resultados da revisão. Higgins e Green (2008) acrescentam que muitas revisões sistemáticas contêm meta-análises, isto é, o uso de métodos estatísticos para resumir resultados de diferentes estudos, destacando que se pode assim combinar todas as informações de pesquisa relevantes para estimar de forma mais precisa efeitos de saúde. Adota-se, a partir de Petticrew e Roberts (2006), a noção de revisão sistemática para revisões que contenham uma análise estatística, isto é, uma meta-análise, e para as que não contenham.

Na busca por desenvolver ferramentas que auxiliem o homem a analisar, interpretar e relacionar dados de forma automática e inteligente, surge a área de Descoberta de Conhecimento em Bases de Dados, ou *Knowledge Discovery in Databases* - KDD (GOLDSCHMIDT; PASSOS, 2005). A mineração de dados consta como uma parte integral da KDD (FAYYAD *et al.*, 1996), sendo definida como o “processo geral de conversão de dados brutos em informações úteis” (TAN; STEINBACH; KUMAR, 2006, p. 4).

Algumas tarefas de mineração de dados são descritas a seguir. A sumarização busca descrever de forma compacta um subconjunto de dados, como no caso de regras de associação e do uso de técnicas de visualização de múltiplas variáveis (FAYYAD *et al.*, 1996). A visualização é uma exibição de dados na forma de gráfico ou tabela, visando à interpretação da informação por uma pessoa ou buscando padrões, o que pode desempenhar papel chave na análise e mineração de dados (TAN; STEINBACH; KUMAR, 2006). A análise associativa é uma ferramenta para descobrir relações interessantes em grandes conjuntos de dados; esses podem ser representados por regras de associação, que têm sua força medida por meio de suporte e confiança. O suporte determina a frequência na qual uma regra é aplicável a um determinado conjunto de dados e a confiança aponta a frequência na qual, a partir de certa premissa, pode ser obtida uma determinada conclusão – por exemplo, em itens que são comprados juntos em um supermercado (TAN; STEINBACH; KUMAR, 2006, p. 392), pode-se pensar numa regra que estabeleça que a partir de que o item *pão* (premissa) e o item *leite* (conclusão) estão presentes em 75% de todas as compras feitas (suporte), o item *leite* estará também presente em 50% das compras com o item *pão* (confiança).

Já a mineração de textos, segundo Mali e Atique (2014), é uma tecnologia para descobrir, de forma semiautomática, padrões, tendências e conhecimentos previamente desconhecidos em bases de texto não-estruturado. Logo, diferencia-se da mineração de dados, que aborda, principalmente, informações já estruturadas. A maioria das informações a serem abordadas na mineração de texto não estão estruturadas, como no caso de textos completos de artigos científicos. Quanto à relação entre mineração de textos e revisão sistemática, tomando a primeira como um suporte à segunda, foi considerada como relativamente desconhecida na comunidade de revisões sistemáticas, bem como necessitada de maior avaliação e desenvolvimento de métodos para compreendê-la (THOMAS; MCNAUGHT; ANANIADOU, 2011).

Mais especificamente, Jusoh e Alfawareh (2012) definem que os métodos fundamentais para a mineração de textos são o Processamento de Linguagem Natural e a Extração de Informação. O Processamento de Linguagem Natural, segundo os autores, visa a geração e a compreensão da língua natural, como em sistemas de tradução automática ou na computação de representações de significados, considerando análises morfológicas, sintáticas e semânticas. A Extração de Informação busca extrair informações úteis do texto, especificando entidades, eventos e relacionamentos entre fontes e estruturando uma representação que aborde as suas principais informações; tal representação, armazenada em uma base de dados estruturada, pode ser fonte para uma mineração de dados (JUSOH; ALFAWAREH, 2012).

Lahlou (2001) traz uma categorização dos principais tipos de resultados da mineração de textos, o que serve como parâmetro de avaliação, sendo esta: (1) triviais, ou seja, resultados óbvios a ponto de serem desinteressantes; (2) clássicos, sendo consistentes com pesquisas prévias; (3) inesperados ou novas descobertas que o analista pode confirmar a partir de outras fontes ou dados; (4) artefatos, encontrados devido a questões técnicas no processamento de dados; e (5) resíduos, que não podem ser interpretados pelo analista. Por fim, considera-se que Lahlou (2001) postula, sobre a mineração de textos, que o software não é o método, e que os dados de saída fornecidos pelo software não são a análise. Coloca, assim, a relevância da interpretação no processo, sendo esta um processo abduutivo que depende do conhecimento externo do analista sobre a linguagem, o tópico e o software. Estes são, portanto, alguns fatores limitadores, podendo-se concluir que o software é um instrumento para exploração e auxílio na interpretação.

Além disso, pondera-se que a mineração de dados não é um recurso novo nem infalível e depende de uma série de fatores, principalmente de uma noção da questão a ser respondida, dos meios para isso e de uma limpeza dos dados. Larose (2005) considera que é uma técnica fácil de ser mal executada, e

acrescenta que a mineração de dados é que deve se adaptar ao processo humano de resolução de problemas, e não que os humanos devam se moldar a ela, isto é, pode-se considerar que é uma ferramenta ou parte do processo de pesquisa, e não a metodologia ou a pesquisa em si. Destaca-se, também, que abordagens metodológicas relacionadas à mineração de dados são recentes e podem trazer mais recursos de análise e resultados potencialmente relevantes. Isto posto, é apresentada, a seguir, a metodologia do estudo.

3 METODOLOGIA

O presente trabalho caracteriza-se como uma pesquisa quantitativa de alcance descritivo (SAMPIERI; COLLADO; LUCIO, 2013). Tendo como ponto de partida o objetivo de identificar e encontrar relações presentes em tópicos de pesquisas sobre Design, esse estudo divide-se em dois tipos de tecnologias de análise de dados, ambas apoiadas por uma análise da amostra com estatística descritiva: a primeira, a partir da mineração de textos, busca compreender os principais tópicos de pesquisa no *corpus*; a segunda, a partir da mineração de dados, busca regras e associações entre os tópicos, por meio das palavras-chave dos artigos e dos conceitos e palavras-chave resultantes da mineração de texto.

O periódico *Design Studies* foi escolhido como foco de estudo tendo em vista que abarca diversos domínios relacionados à área do Design e áreas afins (CROSS, 2007). Além disso, em estudos anteriores, foi julgado por editores e membros de comitês editoriais como um dos melhores periódicos em Design (GEMSER *et al.*, 2012). Não obstante, *Design Studies* é um dos primeiros (CROSS, 2007), mais antigos e mais citados na área (CHAI; XIAO, 2012) e encontra-se disponível digitalmente desde 2001, condição que viabilizou operacionalmente a realização da pesquisa. Sua política editorial está disponível on-line (ELSEVIER, c2017b). Segundo a base de dados Scopus (SCOPUS, 2017), em 2016, o *Design Studies* teve um CiteScore (ELSEVIER, c2017a), índice que mede o número médio de citações recebidas nos três anos anteriores, bem posicionado, ainda mais frente a outras revistas científicas no Design. O periódico está localizado no ranking de CiteScore em 1º de 87 periódicos na área de Arquitetura, 3º de 201 em Ciências Sociais, 9º de 246 em Artes e Humanidades (miscelânea), 13º de 265 em Engenharia, 28º de 153 em Inteligência Artificial e 66º de 511 em Aplicações de Ciência da Computação.

Os artigos selecionados para análise neste estudo estão contemplados pela premissa de que o formato digital configura-se como prioridade na atual produção científica (RIBEIRO *et al.*, 2013). Isto posto, a amostra escolhida para análise abrange artigos publicados no periódico *Design Studies* entre os anos de 2001 e 2016. Esse período foi delimitado devido a intenção de compreender tópicos da Pesquisa em Design na sua atualidade e, ao mesmo tempo, abordar um número significativo de publicações, a fim de se obter robustez de dados.

Cabe ressaltar que esse estudo, sendo parte de uma pesquisa maior, em que se analisou diversas variáveis, aponta resultados relacionados apenas a tópicos de pesquisa, abarcados pelas palavras-chave dos artigos e pelos conceitos e palavras-chave retornados pela mineração de textos. Na Figura 1, é delineado o percurso metodológico do estudo.



FIGURA 1
Percurso metodológico do estudo
Fonte: Elaborada pelos autores.

Após a seleção da fonte de dados, foram delineadas escolhas metodológicas buscando uma filtragem da amostra. Nessa filtragem, alguns artigos foram descartados, visto que não constavam como Artigos de Pesquisa Original (*Original Research Article*), mas sim como artigos de discussão (*Discussion*), Comunicações Curtas (*Short Communication*), Estudos de Caso (*Case Study*) ou artigos de Ponto de Vista (*Viewpoint*), ou seja, não eram artigos completos de pesquisa. Essa escolha deu-se devido à avaliação de que não se adequavam ao escopo do estudo, o qual foca em artigos completos de pesquisa científica em Design. Assim, também foram desconsiderados editoriais do periódico, resenhas e todos os outros tipos de documentos publicados, como Calendário, Pareceristas, Corpo Editorial, Prêmios, Conferências, Discussões, Errata. A especificação dessa filtragem pode ser visualizada na Tabela 1.

TABELA 1
Número de documentos filtrados de todas as publicações do periódico
Design Studies (2001-2016), concluindo uma amostra final de 483 artigos.

Número total de documentos publicados no período	719
Artigos de discussão	1
Short Communication	1
Pontos de vista	6
Estudos de caso	4
Resenhas	22
Editoriais	24
Outros	178
Amostra final	483

Fonte: Elaborada pelos autores

Portanto, a definição da amostra apresenta o número final de 483 artigos do periódico Design Studies, publicados de 2001 a 2016. Para fins de catalogação do *corpus* e de fundamentação para o tratamento e mineração de dados, foi feita uma planilha com uma tabulação dos 483 artigos selecionados, elencando individualmente as variáveis pesquisadas. Para uma caracterização da amostra a partir de estatística descritiva, foram formados subconjuntos de dados, sendo que, nesse estudo, trabalhou-se a partir do subconjunto de palavras-chave. A partir dos subconjuntos, foram construídos gráficos a fim de estabelecer a visualização dos dados analisados e foi aplicada a técnica de análise associativa.

No que tange os dados disponibilizados, tem-se como ponto de partida o fato de que não podem ser usados diretamente devido a discrepâncias encontradas, como diferentes escritas para um mesmo termo ou

a possibilidade de erros de digitação. Para corrigir tais questões, é usado o que se chama de limpeza dos dados (TAN; STEINBACH; KUMAR, 2006). Portanto, as variáveis foram higienizadas, a fim de uma padronização dos dados. Nessa lógica, as palavras-chave foram analisadas de forma a unificar diferentes terminologias para uma mesma ideia, como “*user behavior*” e “*user behaviour*”, ou palavras no singular e plural como “*design tool*” e “*design tools*”. As que foram compreendidas como oriundas de um mesmo tópico foram padronizadas, tornando-se, ainda no exemplo anterior, “*design tool(s)*”. Tal limpeza de dados serviu para auxiliar a aplicação de estatística descritiva e a construção de gráficos para visualização e caracterização da amostra, além de ajustar os dados para a tarefa de análise associativa.

Como tecnologias de apoio ao estudo, foram usados: (1) a AlchemyAPI, para a mineração de textos, focada em extração de conceitos e de palavras-chave; (2) o RapidMiner Studio, para a análise associativa; e (3) softwares de apoio, como Excel, Google Planilhas e Notepad++. A AlchemyLanguage é uma coleção de APIs[1] de Processamento de Linguagem Natural (IBM, c2017b) que faz parte do IBM Watson – plataforma de inteligência artificial que engloba uma série de produtos, serviços e ferramentas direcionadas à utilização de dados para tarefas em diversas áreas. As APIs do AlchemyLanguage contêm funções de análise e mineração de textos que extraem informações semânticas dos dados analisados (IBM, c2017a), acrescentando recursos baseados em aprendizado de máquina, já pré-treinados, a diversas tarefas computacionais (IBM, c2017b). Nesse sentido, é mencionada como fonte para os modelos pré-treinados da ferramenta a Wikipedia. O retorno oferecido pela AlchemyAPI considerado como mais relevante para este estudo é a identificação de conceitos, isto é, termos que a ferramenta detecta como associados ao texto enviado como *input*, a partir de outros conceitos e entidades relacionadas que não estão presentes no texto, advindos de bancos de dados de páginas da web. Dessa forma, podem ser apontados conceitos que não se encontram referenciados no texto, possibilitando uma análise em nível mais complexo do que apenas a identificação de palavras-chave. A função de identificação de conceitos pode ser encontrada como uma ferramenta de visualização dos principais tópicos de um artigo em bibliotecas digitais como a ACM (ACM, 2017), o que colabora para sua relevância como sumarização de produção científica.

Já o RapidMiner Studio é uma plataforma de ciência de dados que integra em um formato visual diversas utilizações, como a extração e preparação de dados, o aprendizado de máquina, mineração de textos e análises preditivas (RAPIDMINER, c2017). É usado em diversas áreas, como negócios, pesquisa, educação, treinamento, prototipagem rápida e desenvolvimento de aplicações (HOFMANN; KLINKENBERG, 2013). O software oferece ferramentas relacionadas às tarefas descritivas de mineração de dados, como a análise associativa, que interessam a este estudo.

Dessa forma, os softwares foram escolhidos tendo em conta sua disponibilidade gratuita, o que viabiliza a operacionalização da investigação, bem como sua robustez em termos de resultados possíveis, a fim de responder aos objetivos da pesquisa. No caso da AlchemyAPI, foi levado em conta que permitia uma mineração de textos a partir dos escritos completos dos artigos, já possuindo uma função pré-estabelecida para tanto, e com velocidade e precisão já apontadas (RIZZO; TRONCY, 2011). Quanto ao RapidMiner Studio, suas funções adequavam-se para execução tanto de acordo com as planilhas iniciais quanto com as planilhas de resultados da AlchemyAPI, além de ser conceituada como uma ferramenta popular e madura de mineração de dados (BATHLA; KATHURIA, 2015). Por fim, os outros softwares serviram de suporte para a base de dados por meio de planilha, sendo ferramentas para a seleção de subconjuntos de dados e construção de gráficos e visualizações, e o Notepad++ foi usado para operações de limpeza, como localização e substituição de dados e expressões regulares.

Como procedimento para a mineração de texto das composições completas dos artigos analisados, seus arquivos foram todos convertidos para a extensão TXT e foram excluídas todas as seções pré-textuais, como título, resumo e palavras-chave, e pós-textuais, como agradecimentos, referências e notas. A AlchemyAPI foi integrada a um software escrito na linguagem Python, criado para executar a mineração. Os parâmetros de configuração foram delineados a partir de um arquivo JSON[2], definindo um mínimo de cinco ocorrências

para um conceito ou palavra-chave ser avaliado como relevante. Também se definiu termos a serem ignorados, como os presentes em cabeçalhos e rodapés, relacionados ao nome do próprio periódico e sua editora. Como retorno da mineração, o software cria dois arquivos CSV[3] para cada um dos retornos (conceitos ou palavras-chave), considerando do mais ao menos relevante: (1) total de termos; (2) e o total de termos por arquivo.

Já para a execução da análise associativa, a partir da mineração de textos, foram utilizadas planilhas binárias. Assim, pode ser analisada a ocorrência conjunta de elementos como, por exemplo, se é possível encontrar uma regra que determine a recorrência da associação entre as palavras-chave *engineering design* e *case study*. No presente trabalho, as regras serão tratadas a partir de seu suporte e sua confiança. Para realizar o processo de análise associativa no software RapidMiner por meio do algoritmo FP-Growth (RAPIDMINER, 2018), foi utilizada uma planilha com as variáveis encontradas, divididas por coluna, em relação aos artigos, divididos por linha. A planilha continha uma relação binária: constava o valor zero quando o termo não era encontrado no artigo, e o valor um quando o termo era encontrado no artigo. Essa transformação de texto não-estruturado formatando vários documentos em uma planilha com a ocorrência de cada termo da coleção em cada documento é chamada, na literatura de Recuperação de Informações, de indexação, sendo o índice, neste caso, uma matriz de incidências (MANNING; RAGHAVAN; SCHÜTZE, 2009, p. 3). Exposta a metodologia, são apresentados a seguir os resultados.

4 RESULTADOS E DISCUSSÃO

Tendo em vista o processo de limpeza dos dados referido na Metodologia, que integrou em apenas uma palavra-chave diversos usos de termos, foram identificadas 458 palavras-chave únicas, ou seja, sem considerar suas repetições em vários artigos. Ademais, houve um número de 2168 palavras-chave no total, com uma média de 4,5 por artigo. Para a Figura 2, foram consideradas apenas as palavras-chave que constaram em 20 (4%) ou mais publicações.

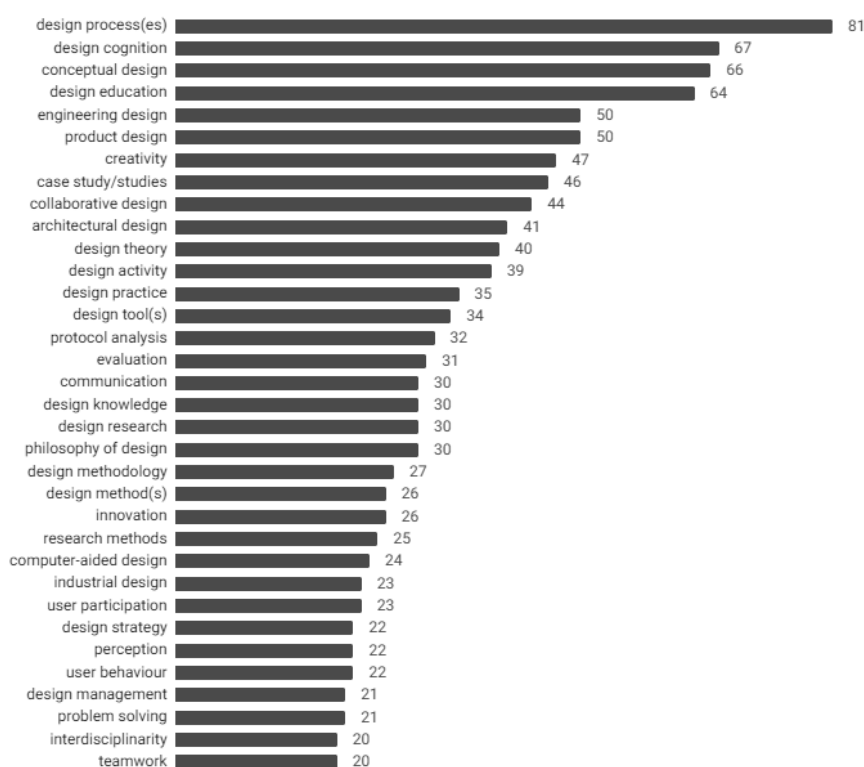


FIGURA 2

As palavras-chave mais citadas em publicações no Design Studies (2001-2016)

Fonte: Elaborada pelos autores.

As áreas de processos e cognição em Design, já apontadas por Chai e Xiao (2012) como predominantes no periódico, são as mais citadas nas palavras-chave; também citado pelos autores, o método da análise de protocolos faz-se um dos mais citados. Além deste, poucas palavras-chave referem-se a métodos de pesquisa, como estudos de caso e avaliações em geral. Também são citadas possíveis áreas de muitos artigos teóricos como Teoria, Filosofia, Pesquisa e Conhecimento em Design.

Além da preocupação com o processo de design ser evidenciada na palavra-chave mais recorrente, podem ser observadas ainda diversas outras relacionadas, que a reforçam como postura predominante no periódico: “*design activity/practice/tool(s)/methodology/method(s)/strategy*”, “*computer aided design*”, “*problem solving*” e “*design management*”, a qual já havia sido indicada como em ascensão no periódico por Ribeiro e colaboradores (2013). Nessa perspectiva, também se percebe a preocupação com a colaboração durante o projeto, como em “*collaborative design*”, “*user participation*” e “*teamwork*”, e com a criação, como em “*conceptual design*” e “*creativity*”, além de “*innovation*”. Nota-se também nas palavras-chave as diversas áreas abarcadas pelo periódico (ELSEVIER, c2017b), o que indica a integração entre áreas do conhecimento, já apontada por Friedman (2000) como característica do Design. Outros tópicos indicam uma possível preocupação com usuários: “*perception*” e “*user behaviour*”. Tomando esses focos como principais, pode-se sugerir como direcionamento geral o foco no processo de design em diversas áreas, buscando principalmente a etapa de criação de conceitos, tanto pelo viés da cognição – processos cognitivos individuais de designers nos projetos – quanto pelo viés da colaboração, em equipes de designers ou junto a usuários ou clientes.

Além dos resultados concernentes às palavras-chave dos artigos, apresenta-se aqui os resultados de mineração de textos obtidos a partir da contagem e cálculo de relevância da ferramenta AlchemyAPI a partir

dos textos completos dos artigos do *corpus*. Inicialmente, são abordadas as palavras-chave identificadas pela ferramenta, na Tabela 2.

TABELA 2
Palavras-chave apontadas pela ferramenta AlchemyAPI como mais relevantes nos textos completos de publicações no Design Studies (2001-2016), com quantidade e relevância

Palavras-chave	Qt.	Rel.	Palavras-chave	Qt.	Rel.
Design	228	0,90	design research	6	0,65
conceptual design	6	0,89	Designers	14	0,62
project	6	0,86	Problem	10	0,62
product	12	0,83	Ideas	9	0,62
idea generation	13	0,81	engineering design	9	0,55
design activity	6	0,78	design task	6	0,49
students	13	0,74	design problem	8	0,48
information	6	0,71	team members	6	0,45
design process	45	0,69	0142-694X Design Studies	10	0,40
concepts	7	0,67			

Fonte: Elaborada pelos autores com base nos dados retornados pela AlchemyAPI (IBM, c2017b) em 2018.

A extração de palavras-chave pela ferramenta AlchemyAPI se baseia principalmente nas palavras usadas no *corpus* e na relação entre elas, formando uma seleção do que é mais citado. Reitera-se que estas palavras-chave são diferentes das apresentadas na Figura 2, as quais constavam como metadados dos próprios artigos – já as palavras-chave apresentadas nesta Tabela 2 foram detectadas pela AlchemyAPI. Disso, se considera que a quantidade, na tabela, significa o número de artigos que a ferramenta apontou que apresentavam certo termo como palavra-chave a partir de seu texto completo. Inicialmente, a palavra-chave “*Design*” é trivial (LAHLOU, 2001), visto sua evidente relevância no periódico, que é dessa área. Percebeu-se, também, que a ferramenta tende a apontar mais palavras-chave simples – de um termo só – do que compostas, como era a maioria nos metadados dos artigos. Nessa lógica, percebeu-se que, à exceção de “*Design*”, algumas palavras-chave de apenas um termo identificadas pelo AlchemyAPI parecem se relacionar a um caráter mais específico do procedimento de pesquisas, como “*designers*”, “*students*” e “*concepts*”, tendo um aspecto mais prático e específico do que teórico e geral, e podem referir-se a poucos artigos, visto que um artigo pode citar muitas vezes termos relacionados ao seu procedimento. Ainda nos termos citados, cogita-se que estudos conduzidos com universitários são facilitados pelo caráter da pesquisa acadêmica na área, realizada em grande parte por pesquisadores que são professores universitários e que, portanto, podem contar com os estudantes como sujeitos de pesquisa.

Nota-se, ainda, que as palavras-chave identificadas pela AlchemyAPI não apontaram para tantas áreas de estudo quanto as palavras-chave dos artigos. O tópico “*design management*”, previamente apontado como em ascensão (RIBEIRO et al., 2013), não foi apontado como palavra-chave, o que gera questionamentos em relação ao uso efetivo das palavras-chave dos artigos em seus textos, e se servem essas como um parâmetro para o que está no texto ou se acabam sendo mais somente um enquadramento teórico-metodológico. O tópico do processo de design, o mais relevante segundo as palavras-chave dos artigos, mostrou-se como a palavra-chave mais identificada, além de outras que podem ser avaliadas como correlatas (“*design process/activity/activities/project*”). Também pode-se ver uma abordagem relacionada a problemas ou tarefas de design, o que sugere que foram conduzidas pesquisas com designers e/ou estudantes que deveriam solucionar essas tarefas. Já o tópico da cognição, qualificado como relevante nos resultados anteriores, não teve destaque, assim como o foco em métodos de pesquisa.

Não houve ênfase na abordagem de questões teórico-epistemológicas em relação à Ciência do Design ou relação do Design com a ciência, assim como em relação à Filosofia do Design, em destaque segundo

as palavras-chave dos artigos. Permanece a questão para futuros estudos quanto a abordagem da Pesquisa em Design como uma disciplina (OWEN, 1998), em busca dos significados imputados à palavra-chave “*design research*” – por exemplo, se tratam como uma disciplina ou um método. Já o tópico da Educação em Design teve destaque aqui por meio da palavra-chave “*student(s)*”, ainda não abordada, além de, novamente, “*design education*”, o que corrobora com autores em relação ao crescimento da pós-graduação na área e do enfrentamento de desafios educacionais (CROSS, 2007; LLOYD, 2017; BAYAZIT, 2004), mostrando um amadurecimento da área.

Algumas questões identificadas por meio de diversas palavras-chave referem-se ao design feito em conjunto ou em equipes, o que pode ser julgado como um tópico relevante junto aos metadados em relação à “*collaborative design*”. Outras questões, que podem sugerir campos específicos de pesquisas, referem-se ao espaço onde se executam os projetos de design e as diversas questões relacionadas às propriedades de produtos (“*product(s)*”, “*function(s)*”, “*shape*”, “*artefact*”, “*product family*”). Há também uma diversidade de palavras-chave associadas à criatividade (“*idea generation*”, “*ideas*”, “*concepts*”, “*conceptual design*”, “*creativity*”, “*novelty*”, “*design concept*”), o que consolida esse como um dos principais tópicos abordados no periódico. Acredita-se que as palavras-chave identificadas a partir dos textos completos dos artigos podem trazer diferenças quanto às palavras-chave de cada artigo, levantando questões sobre a qualidade do retorno da ferramenta e sobre qual âmbito – a partir dos metadados ou dos textos completos – pode gerar resultados mais relevantes. Sendo assim, comparou-se as 20 palavras-chave mais citadas nos metadados (Figura 2) com as mais relevantes nos textos completos, segundo a AlchemyAPI (Tabela 2).

Foram cinco as palavras-chave presentes em ambos os âmbitos. Dessas, “*design activity*” foi duplicada no AlchemyAPI em “*design activities*”, o que, junto a sua alta frequência de citação, pode corroborar para sua relevância no periódico. Pondera-se também que algumas palavras-chave servem mais para um enquadramento epistemológico-metodológico da publicação, o que as levou a não serem muito citadas nos textos. Exemplos: “*case study/studies*”, “*architectural design*”, “*protocol analysis*”, “*philosophy of design*”. Por último, mesmo que grande parte das palavras-chave dos dois âmbitos sejam diferentes entre si, diversas delas estão relacionadas, como “*creativity*”, que pode ser associada a “*idea generation*”, “*ideas*” e “*concepts*”; “*design education*”, que se relaciona a “*students*”; “*collaborative design*” e “*team members*”; “*product design*” e “*product*”. Algumas palavras-chave que constavam nos metadados também foram identificadas pelo AlchemyAPI, apesar de que com menos frequência: “*product design*”, “*creativity*”, “*design education/practice*”. Assim, pode-se dizer que a mineração dos textos completos feita pela ferramenta é válida e se relaciona com as palavras-chave dos artigos, ainda levantando novos tópicos. Além das palavras-chave, são apresentados na Tabela 3 os conceitos retornados pela AlchemyAPI.

TABELA 3

Os conceitos apontados pela ferramenta AlchemyAPI como mais relevantes nos textos completos de publicações no Design Studies (2001-2016), com quantidade e relevância

Conceitos	Qt.	Rel.	Conceitos	Qt.	Rel.
Design	408	0,90	Experience design	33	0,51
Cognition	29	0,69	Communication design	38	0,50
Problem solving	45	0,67	Interaction design	35	0,50
Scientific method	26	0,64	Design management	189	0,48
Psychology	24	0,64	User-centered design	21	0,48
Creativity	81	0,63	Design research	121	0,47
Knowledge	24	0,63	Engineering design process	82	0,47
Architecture	22	0,60	Designer	80	0,46
Industrial design	60	0,56	Architectural design	49	0,45
Graphic design	140	0,54	Design strategy	40	0,45
Engineering	102	0,53	Design thinking	28	0,45
Mechanical engineering	27	0,52	Design engineer	31	0,44
Software design	24	0,52	Design education	22	0,40

Fonte: Elaborada pelos autores com base nos dados retornados pela AlchemyAPI (IBM, c2017b) em 2018.

Novamente, segundo Lahlou (2001), o conceito “*Design*” pode ser classificado como trivial. Na categoria de resultados clássicos, reforçando a literatura já existente, percebeu-se que a maioria dos conceitos encontrados se refere às mais diversas áreas de estudo dentro do Design e correlatas. Isso vai ao encontro da interdisciplinaridade (FRIEDMAN, 2001), mesmo que essa não tenha sido apontada como conceito, e pode dar cor ao argumento do crescimento do arsenal teórico do Design como campo (BECCARI, 2012). Aliás, a área de “*design management*”, aqui apontada como conceito recorrente, já havia sido indicada como em ascensão no periódico por Ribeiro e colaboradores (2013); e o conceito “*project management*” também pode estar relacionado.

Observou-se inesperada recorrência do conceito de “*Scientific method*”, apesar de nenhuma outra recorrência relacionada a métodos, além de “*Case study*” - que aqui não esteve em evidência. É recorrente também o conceito “*Knowledge*”, mas não se pode afirmar que se refira à pesquisa; pode ser voltado, nos artigos, ao processo e à Educação em Design – algo sustentado pela possível relação com outros conceitos que abordam educação –, o que suscita outras pesquisas, também a se verificar a ligação com as perspectivas dos autores quanto ao conhecimento científico e ao Design como área de conhecimento. Não pôde ser verificada, a partir dos conceitos encontrados, a recorrência da análise de protocolos como um método no periódico, a qual já havia sido apontada por Chai e Xiao (2012) e corroborada pelos metadados. Em outro âmbito, constaram nos resultados áreas não tão comumente relacionadas, como a Psicologia e a Sociologia, além da Engenharia Mecânica, que pode evidenciar, junto a “*Design engineer*”, um amadurecimento da relação entre Design e Engenharia, já em ascensão desde a década de 1980 (BAYAZIT, 2004; CROSS, 2007). Algumas questões de caráter mais teórico-epistemológico apontadas pelos autores quanto à relação do Design com a ciência e a Filosofia do Design não foram contempladas com destaque por nenhum conceito, ao contrário dos metadados. Já a Pesquisa em Design aparece nos conceitos, assim como nos metadados, como um dos tópicos de maior destaque. Logo, sua recorrência motiva mais estudos para compreender quais os significados imputados a “*Design research*”, visto as divergências teóricas nessa questão. Pode-se considerar que a ascensão do tópico da Educação no Design, evidenciada pelos conceitos “*Design Education*”, “*Educational psychology*” e “*Education*” e, nos metadados, por “*design education*”, aponta para um amadurecimento que já havia sido

observado pelos autores em relação ao crescimento da pós-graduação na área (CROSS, 2007), do número de PhDs (LLOYD, 2017) e dos desafios educacionais (BAYAZIT, 2004). Crê-se, na esteira de Lloyd (2017), que tal foco na Educação e na Pesquisa em Design, após 50 anos (OWEN, 1998), mostra popularização e fortalecimento da área. Outro tópico que se mostrou relevante no *corpus*, tanto nas palavras-chave dos artigos quanto nos conceitos da mineração de texto, é o da criatividade. De certa forma, o crescimento dessa relevância na atualidade é curioso, visto que é nos anos 1950 e 1960 que há um contexto de necessidade de novos métodos de criatividade (CROSS, 2007; BAYAZIT, 2004) e são lançados os primeiros livros sobre técnicas de criatividade (BECCARI, 2012). Também, por ainda não ter sido apontada pelos autores como um tópico em destaque, se mostra um achado relevante do estudo, apontando para o foco recente nas etapas de geração de ideias no design. Isso pode estar relacionado a um contexto de dificuldades para lidar com a grande quantidade de informação no processo criativo – alguns estudos abordam a oposição entre “inspiração” e “fixação” – ou à necessidade de produtividade para gerar mais e melhores ideias e, ainda, à busca por compreender em termos cognitivos o processo criativo do designer. Por outro lado, ressalta-se que o tópico da cognição não é tão incidente aqui quanto nas palavras-chave dos trabalhos, o que parece relevante, visto que é um dos tópicos destacados na página do periódico e apontado em estudo prévio como predominante (CHAI; XIAO, 2012). Uma possível explicação é de que alguns estudos que apontam a cognição nas palavra-chave usem-na mais como enquadramento teórico-epistemológico, e não necessariamente a abordem no texto, focando mais em outros conceitos. O mesmo se pode sugerir a respeito da palavra-chave mais predominante, “*design process(es)*”, que aqui não apresentou tanta incidência – pode servir como palavra-chave, mas não ser tão citada no texto, que trata mais dos detalhes do processo a serem estudados.

Para concluir a análise relacionada à mineração de texto, pode-se inferir que a ferramenta AlchemyAPI oferece substanciais resultados para identificação de tópicos de artigos científicos a partir de seus textos completos, tão válidos quanto as próprias palavras-chave dos artigos. Os resultados, em relação aos conceitos, foram observados como mais específicos e com o potencial de levantar novas questões, tendo em vista que também apontam conceitos que não estão elencados no texto. Quanto às palavras-chave identificadas pela ferramenta, vê-se que têm caráter mais prático e específico, enquanto as palavras-chave dos artigos são mais teóricas e gerais, provavelmente pela necessidade de suprir a demanda de um enquadramento epistemológico e metodológico abrangente. A seguir, são abordadas as regras de associação entre as palavras-chave dos metadados dos artigos (Tabela 4)[4]:

TABELA 4
Regras de associação entre palavras-chave do Design Studies (2001-2016)

Premissa	Conclusão	Suporte	Confiança
design science	design research	1,2%	60%
case study/studies, design activity	design practice	1,0%	83%
case study/studies, design practice	design activity	1,0%	71%
design activity, design practice	case study/studies	1,0%	71%
Styling	aesthetics	1,0%	56%
learning style(s)	design education	0,8%	100%
design fixation	creativity	0,8%	100%
urban design	design education	0,8%	67%
Affordance	product design	0,8%	67%
interior design	design education	0,6%	100%
service design	case study/studies	0,6%	100%
experiment	research methods	0,6%	100%
collaborative design, Creative design	design cognition	0,6%	100%
Creative design, design technique(s)	conceptual design	0,6%	100%
user behaviour, perception	product design	0,6%	100%
product design, styling	aesthetics	0,6%	100%
collaborative design, design management	case study/studies	0,6%	100%
philosophy of design, design	design theory	0,6%	100%

Fonte: Elaborada pelos autores com base nos dados retornados pelo software RapidMiner Studio (2017) em 2018.

Podemos pensar, a partir da segunda regra, como exemplo para melhor compreensão, que, considerando os 483 documentos analisados, 5 destes (1% de suporte) apresentaram a combinação das palavras-chave *case study/studies*, *design activity* e *design practice*. Considerando todos os documentos que apresentaram a premissa *case study/studies* e *design activity*, 83% deles também apresentaram a conclusão *design practice*. Ou seja, haviam 6 documentos com a premissa da regra e destes 6, apenas 5 (83%) apresentam também a conclusão.

A Ciência do Design foi abordada tanto em relação à Pesquisa em Design quanto a métodos de pesquisa e Teoria e Filosofia do Design, o que sugere uma preocupação quanto à relação do Design com a ciência e com seus métodos de pesquisa, além de, por sua relação com a Filosofia, uma tentativa epistemológica de articulação do campo. Destaca-se as palavras-chave que apontam estudos de caso relacionados a atividade e prática do design, bem como design de serviços, design colaborativo e gestão de design, reiterando a predominância do método no periódico. As associações entre “*styling*”, “*aesthetics*”, “*product design*”, “*user behaviour*” e “*perception*” mostram um foco nas questões estéticas de produtos, algo ainda não evidenciado nos resultados.

Observa-se também que a Educação em Design abarcou o foco nos estilos de aprendizagem, bem como em duas áreas específicas não recorrentes, ligadas à Arquitetura: “*urban/interior design*”. Pôde ser observada, também, a conexão da criatividade, tópico já apontado como relevante no *corpus*, com a fixação, com o viés cognitivo e, ainda, com técnicas de design. Isso aponta, novamente, para a busca do que seria um

aprimoramento das habilidades de um designer, compreendendo como funciona o processo criativo em busca de geração maior de ideias, evitando entraves como a fixação. Na Tabela 5, vê-se as coocorrências mais frequentes de palavras-chave.

TABELA 5
Coocorrências de palavras-chave do Design Studies (2001-2016)

Item 1	Item 2	Suporte
design cognition	creativity	3,3%
design theory	philosophy of design	3,3%
design cognition	conceptual design	3,1%
design cognition	protocol analysis	2,5%
design process(es)	design education	2,3%
conceptual design	creativity	2,3%
conceptual design	engineering design	2,1%
collaborative design	communication	2,1%
design process(es)	design cognition	1,9%
design process(es)	architectural design	1,9%
design cognition	collaborative design	1,9%
conceptual design	design activity	1,9%
conceptual design	protocol analysis	1,9%
design education	engineering design	1,9%
design education	creativity	1,9%
product design	perception	1,9%
design	design	1,7%

Fonte: Elaborada pelos autores com base nos dados retornados pelo software RapidMiner Studio (2017) em 2018.

Vê-se aqui os principais tópicos do periódico refletidos nas coocorrências mais frequentes. A Cognição em Design, já apontada como um dos tópicos mais relevantes, relaciona-se com diversos tópicos, assim como o Processo de Design e a etapa de projeto do Design Conceitual. A análise de protocolos, única palavra-chave associada a uma questão metodológica, apresentou-se novamente relacionada à Cognição em Design e ao Design Conceitual.

Observa-se que a preocupação central circunda o Processo do Design, sobretudo na etapa de criação, e uma compreensão da cognição individual nesse processo, que pode ser constatada a partir da técnica da análise de protocolos. Relaciona-se isso com a Educação em Design e com áreas como a Engenharia e a Arquitetura. Além disso, vê-se também o foco relacionado à Teoria e Filosofia do Design, à Comunicação durante a Colaboração em projetos de design e à percepção de usuários, enquanto interesse do Design de Produtos.

TABELA 6
Regras de associação entre conceitos identificados pela ferramenta AlchemyAPI
no texto completo das publicações no Design Studies (2001-2016)

Premissa	Conclusão	Suporte	Confiança
Communication design	Graphic design	7,0%	90%
Design strategy	Design research	6,0%	73%
Design management, Engineering design process	Engineering	5,6%	69%
Design thinking	Design research	5,2%	89%
Design research, Engineering design process	Engineering	4,8%	72%
Design education	Graphic design	4,6%	100%
Graphic design, Engineering design process	Engineering	4,1%	77%
Fluid dynamics	Graphic design	3,5%	100%
Design management, Communication design	Graphic design	3,5%	90%
Psychology	Cognition	3,5%	71%
Design methods	Design research	3,3%	84%
Design strategy, Design thinking	Design research	3,1%	100%
Architectural design, Design strategy	Design research	3,1%	88%

Fonte: Elaborada pelos autores com base nos dados retornados pelo software RapidMiner Studio (2017) em 2018.

Primeiramente, na Tabela 6, o alto suporte e confiança entre “*Communication design*” e “*Graphic design*” sugere uma forte associação do primeiro com o segundo, provavelmente relacionada com o projeto de peças comunicacionais gráficas. Por essas regras, pode-se concluir que “*Design research*” não foi associado principalmente a questões mais teórico-epistemológicas do campo, mas sim à prática do design.

São apontadas como algumas das principais associações as que envolvem a Engenharia e seu processo de projeto associados à Gestão e à Pesquisa em Design. Além disso, o conceito de Dinâmica de Fluidos se relaciona com o Design Gráfico, algo inesperado que suscita aprofundamentos, com vistas a verificar a efetiva procedência dessa relação. Já o tópico da Educação em Design, previamente relacionado a diversas áreas, aqui se relaciona apenas ao Design Gráfico, o que também provoca questionamentos. Por fim, há também a previamente abordada associação entre Psicologia e Cognição, que colabora para afirmar a interdisciplinaridade (FRIEDMAN, 2001) e o crescimento do arsenal teórico do Design (BECCARI, 2012).

Dessa forma, ressalta-se que a análise de associação entre conceitos gerados pela ferramenta AlchemyAPI, a partir dos textos completos das publicações do *corpus*, possibilitou a verificação de relações não encontradas

apenas com os metadados dos artigos, o que aponta para a sua relevância enquanto uma ferramenta para descoberta de conhecimentos.

5 CONSIDERAÇÕES FINAIS

Este estudo buscou identificar e verificar as relações entre tópicos de pesquisa em publicações de 2001 a 2016 no periódico *Design Studies* empregando revisão sistemática, com o suporte de tecnologias de KDD, por meio das ferramentas *AlchemyAPI* e *RapidMiner Studio*. A amostra analisada foi de 483 artigos, em abordagem quantitativa de alcance descritivo. Inicialmente, foi caracterizada a amostra com estatística descritiva, analisando as palavras-chave das publicações. Posteriormente, foram aplicadas mineração de textos para extração automática de conceitos e palavras-chave dos textos completos dos artigos e análise associativa a partir dos metadados e da mineração de texto.

Concluiu-se que a ferramenta *AlchemyAPI* oferece, para identificação de tópicos de artigos científicos a partir de seus textos completos, resultados tão válidos quanto as próprias palavras-chave dos artigos. Nesse ponto de vista, propõe-se que somente a aplicação de mineração de texto gera por si só resultados relevantes, sem considerar os metadados dos artigos, mesmo que esses sirvam como parâmetros para comparação de resultados em uma abordagem mais completa. Assim, a mineração de textos, citada antes como relativamente desconhecida na comunidade de revisões sistemáticas, constitui-se como um relevante suporte. Neste estudo, os Conceitos da *AlchemyAPI* foram observados com o potencial de levantar novos questionamentos, tendo em vista que também apontam tópicos que não estão efetivamente citados no texto, e as palavras-chave identificadas foram mais práticas e específicas que as dos metadados, as quais são mais teóricas e gerais, possivelmente pela busca por enquadramento epistemológico abrangente para os estudos.

Como limitações do estudo, aponta-se a configuração do recorte, que permite a visualização de dados limitados e somente de uma área. Nesse sentido, considera-se que a aplicação das técnicas de KDD aqui abordada pode ser estendida, em estudos futuros, a outras áreas científicas, como, por exemplo, as áreas relacionadas no periódico ao *Design*. Isso porque a *AlchemyAPI* apresentou potencial para a detecção automática de tópicos de pesquisa em artigos científicos, o que enseja uma abordagem que contemple maior quantidade de dados e possibilite uma comparação com outras áreas do conhecimento, a fim de verificar possíveis padrões, ou mesmo analisar se há áreas em que a técnica é mais aplicável, de acordo com suas especificidades. Além disso, é importante ressaltar que no caso deste estudo foram aplicadas apenas alguns conjuntos de software, mas que a partir dos mesmos algoritmos pode-se reproduzir os mesmos resultados a partir de outras ferramentas.

Há de se elencar o fato de que, conforme Lahlou (2001), é relevante que o pesquisador detenha conhecimento da área na qual se realiza a revisão sistemática, visto que as tecnologias servem, a partir dos conhecimentos descobertos, para sugestão de questionamentos a serem possivelmente destrinchados em outros estudos, reflexão que exige certo conhecimento específico do campo em análise por parte do pesquisador. Outras possíveis tecnologias que poderiam ser exploradas para o mesmo problema de pesquisa são a análise de redes, frequentemente empregada para análise de citações, redes de co-autoria etc. – como, por exemplo, observado em Chai e Xiao (2012) e Gemser e De Bont (2016) –, e agrupamento de artigos segundo determinadas variáveis, os quais não foram foco do presente trabalho.

Espera-se que o presente percurso metodológico (visto na Figura 1) possa ser apropriado a estudos que conduzam aplicações de revisão sistemática de literatura no *Design* e em outras áreas, a fim de que possa ser um caminho frente a necessidade de técnicas que apoiem a investigação de tópicos de pesquisa em publicações científicas. Nesse sentido, mesmo tendo em vista que as técnicas de mineração ainda precisam ser mais exploradas, acredita-se que podem ser um auxílio a trabalhos que se proponham a gerar conhecimento acerca de um campo. Nesse sentido, obter conhecimento e ciência dos tópicos mais abordados a nível global também pode colaborar para aumentar a presença do Brasil em periódicos de relevância internacional na

área do Design, como o Design Studies, favorecendo tanto o intercâmbio de conhecimentos com outros pesquisadores quanto a divulgação e o fomento de produções relevantes no país.

Por fim, estabelece-se que a mineração de texto, por meio da extração automática de conceitos e palavras-chave, pode servir a uma melhor identificação dos principais tópicos em um *corpus*, nem sempre evidentes apenas nos metadados de publicações. Já a mineração de dados, por meio da análise associativa, pode servir para a descoberta de relações mais frequentes entre variáveis com regras de associação e, também, as coocorrências em um *corpus* possibilitam identificar relações, mas sem a formalização de uma regra. A estatística descritiva, por fim, pode servir tanto à análise de palavras-chave como de outras variáveis, como autores, nacionalidades, instituições, coleta de dados e métodos.

AGRADECIMENTOS

Agradecemos à CAPES pela bolsa concedida, que possibilitou a finalização desta pesquisa, originada na dissertação de mestrado do primeiro autor.

REFERÊNCIAS

- ACM. ACM DL Digital Library. 2017.
- AKOBENG, A. K. Understanding systematic reviews and meta-analysis. *Archives of disease in childhood*, London, v. 90, p. 845-848, 2005.
- ALVAREZ, G. R.; CAREGNATO, S. E. A Ciência da Informação e sua contribuição para a avaliação do conhecimento científico. *Biblos: Revista do Instituto de Ciências Humanas e da Informação*, Rio Grande, v.31, n.1, p.9-26, jan./jun. 2017.
- BATHLA, H; KATHURIA, K. Apriori algorithm and filtered associator in association rule mining. *International Journal of Computer Science and Mobile Computing*, Secunderabad, v. 4, n. 6, p. 299-306, jun. 2015.
- BAYAZIT, N. Investigating design: a review of forty years of design research. *Design Issues*, Cambridge, v. 20, n. 1, p. 16–29, 2004.
- BECCARI, M. Articulação simbólica: uma abordagem junguiana aplicada à filosofia do design. 2012. Dissertação (Mestrado em Design) – Universidade Federal do Paraná, Curitiba, 2012.
- BORGES, V. A.; NOGUEIRA, B. M.; BARBOSA, E. F. Uma análise exploratória de tópicos de pesquisa emergentes em Informática na Educação. *Revista Brasileira de Informática na Educação*, Porto Alegre, v. 23, n. 1, p. 1-13, mar. 2015.
- BUCHANAN, R. Design Research and the New Learning. *Design Issues*, Cambridge, v. 17, n. 4, p. 3–23, 2001.
- CHAI, K.-H.; XIAO, X. Understanding design research: A bibliometric analysis of Design Studies (1996–2010). *Design Studies*, [s.l.], v. 33, n. 1, p. 24–43, 2012.
- CROSS, N. Designerly ways of knowing: design discipline versus design science. *Design Issues*, Cambridge, v. 17, n. 3, p. 49–55, 2001.
- CROSS, N. Editorial: Forty years of design research. *Design Studies*, [s.l.], v. 28, n. 1, p. 1–4, 2007.
- ELSEVIER. CiteScore. c2017a.
- ELSEVIER. Design Studies. c2017b. Disponível em: <http://www.journals.elsevier.com/design-studies>. Acesso em 07 jul. 2017.
- FAYYAD, U.M.; PIATETSKY-SHAPIO, G.; SMYTH, P. Knowledge discovery and data mining: towards a unifying framework. In: *INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING*, 2., 1996, Portland.. *Proceedings [...]* Portland, Oregon, 1, p. 33-47, 1996.
- FRIEDMAN, K. Creating design knowledge: from research into practice. In: *IDATER 2000, Proceedings [...]* Loughborough University, p. 05-32, jun. 2001.

- 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, v. 2, n. 1, p. 46–58, 2016.
- GEMSER, G. et al. Quality perceptions of design journals: The design scholars' perspective. *Design Studies*, [s.l.], v. 33, n. 1, p. 4–23, 2012.
- GOLDSCHMIDT, R.; PASSOS, El. Data mining: um guia prático. Conceitos, técnicas, ferramentas, orientações e aplicações. Rio de Janeiro: Campus, 2005.
- HIGGINS, J. P. T.; GREEN, S. (Eds.). *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: John Wiley & Sons Ltd., 2008.
- HOFMANN, M.; KLINKENBERG, R. RapidMiner: data mining use cases and business analytics application. *Chapman & Hall/CRC Data Mining and Knowledge Discovery Series*. CRC Press, 25 out. 2013.
- IBM. Alchemy documentation. c2017a. Disponível em: <https://www.ibm.com/watson/developercloud/doc/index.html>. Acesso em 10 jul. 2017.
- IBM. Alchemy language. c2017b. Disponível em: <https://www.ibm.com/watson/developercloud/alchemy-language.html>. Acesso em 05 jul. 2017.
- JUSOH S.; ALFAWAREH, H. M. Techniques, applications and challenging issue in text mining. *IJCSI International Journal of Computer Science Issues*, v. 9, n. 2, p. 431-436, 2012.
- LAHLOU, S. Text mining methods: an answer to Chartier and Meunier. *Papers on Social Representations*, Lisbon, v. 20 n. 38, p. 1-7, 2001.
- LAROSE, D. T. *Discovering knowledge in data: an introduction to data mining*. New Jersey: Wiley, 2005.
- LLOYD, P. From Design Methods to Future-Focused Thinking: 50 years of design research. *Design Studies*, [s.l.], v. 48, p. A1–A8, 2017.
- LOVE, T. Philosophy of design: a meta- theoretical structure for design theory. *Design Studies*, [s.l.], v. 21, n. 3, p. 293–313, 2000.
- MALI, M.; ATIQUÉ, M. Applications of text classification using text mining. *International Journal of Engineering Trends and Technology*, [s.l.], v. 13, n. 5, p. 209-212, 2014.
- MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press. 2009.
- OWEN, C. L. Design research: building the knowledge base. *Design Studies*, [s.l.], v. 19, p. 9-20, 1998.
- PETTICREW, M.; ROBERTS, H. (Eds.). *Systematic Reviews in the Social Sciences: a practical guide*. Padstow: Blackwell Publishing, 2006.
- RAPIDMINER. RapidMiner Studio. 2017. Disponível em: <http://rapidminer.com/>. Acesso em 08 jul. 2017.
- RAPIDMINER. Documentation. FP-Growth. 2018. Disponível em: <http://rapidminer.com/>. Acesso em 19 set. 2018.
- RIBEIRO, V. G.; SILVEIRA, S. R.; SILVEIRA, A. L. M. da; ATKINSON, R.; ZABADAL, J. R. S. The use of data mining techniques for defining strategies in scientific communication processes in design journals. *Strategic Design Research Journal*, Porto Alegre, v. 6, n. 2, p. 85-94, 2013.
- RIZZO, G.; TRONCY, R. NERD: Evaluating named entity recognition tools in the web of data. In: *WORKSHOP ON WEB SCALE KNOWLEDGE EXTRACTION (WEKEX'11)*, Bonn, Alemanha. Proceedings [...]. Porto Institutional Repository, p. 1-16, out. 2011.
- SAMPIERI, R.; COLLADO, C.; LUCIO, M. *Metodologia de pesquisa*. Porto Alegre: Penso, 2013.
- SCOPUS. *Design Studies*. 2017. Disponível em: <https://www.scopus.com/sourceid/36450>. Acesso em: jun. 2017
- SÉRGIO, M.; SILVA, T.; GONÇALVES, A. Descoberta de conhecimento a partir de informações não estruturadas por meio de técnicas de correlação e associação. *Em Questão*, Porto Alegre, v. 22, n. 2, p. 87-113, maio/ago. 2016.
- SOMASUNDARAM, G.; SHRIVASTAVA, A. *Armazenamento e*
- TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. *Introduction to Data Mining*. Edinburgo: Pearson Addison-Wesley, 2006.

- THOMAS, J.; MCNAUGHT, J.; ANANIADO, S. Applications of text mining within systematic reviews. *Research Synthesis Methods*, [s.l.], v.2, p.1-14, jan. 2011.
- TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. *Introduction to Data Mining*. Edinburgo: Pearson Addison-Wesley, 2006.
- THOMAS, J.; MCNAUGHT, J.; ANANIADO, S. Applications of text mining within systematic reviews. *Research Synthesis Methods*, [s.l.], v.2, p.1-14, jan. 2011.

NOTAS

- [1] API é uma sigla para Application Programming Interface — do inglês, Interface de programação de aplicativos, em tradução livre. De modo geral, é um conjunto de rotinas e padrões estabelecidos por um software para a utilização das suas funcionalidades por aplicativos que não pretendem envolver-se em detalhes da implementação do software, mas apenas usar seus serviços.
- [2] JSON, um acrônimo de JavaScript Object Notation, é um formato de padrão aberto para trocas de dados entre sistemas, usando texto legível a humanos, no formato atributo-valor.
- [3] Comma Separated Values (CSV), ou valores separados por vírgula, no português, é um formato de arquivo que pode conter registros textuais estruturados, separados por algum delimitador, como vírgula ou ponto e vírgula. Os arquivos CSV são interpretados como tabelas por editores de planilha eletrônica como Microsoft Excel, Libre Office Calc, e Google Sheets, e também podem ser lidos como arquivos texto por qualquer software.
- [4] Foi usado como critério para geração de regras a confiança mínima de 50%, alterando o suporte de forma iterativa a fim de verificar quais regras eram geradas pelo software. Assim, neste trabalho, não são apresentadas todas as regras geradas, mas sim apenas as de maior suporte em cada âmbito.