

Método visual para revisão sistemática em Design com base em conceitos da Mineração de Dados

Visual method for systematic review in design based on concepts of Data Mining

*Arina Blum*¹

*Eugenio Andrés Díaz Merino*²

*Giselle Schmidt Alves Díaz Merino*³

Resumo

Para o design, área multidisciplinar, é propício encontrar a maneira adequada para realizar a revisão sistemática, na medida em que, por vezes, os termos de busca trazem vastos resultados, inviáveis de serem analisados como um todo. Neste sentido, técnicas de Mineração de Dados podem ser úteis, pois implicam no uso de ferramentas para realização de extração de informações em meio a grande quantidade de dados. Este artigo apresenta os resultados de um trabalho que objetivou desenvolver e aplicar um método visual para revisão sistemática em design com base em conceitos da Mineração de Dados. A pesquisa aplicada, qualitativa na sua forma de abordagem e descritiva quanto aos objetivos, utilizou um estudo de caso para demonstrar o uso do método proposto. Os resultados apontam para a eficiência do método, tendo sido possível, através dele, localizar e filtrar informações de maneira sistemática e produtiva.

Palavras-Chave: design; mineração de dados; metodologia; pesquisa científica

Abstract

To a multidisciplinary field as design, finding a suitable way to accomplish the systematic review is convenient, since sometimes the search terms bring back vast results, which are impracticable to be analyzed as a whole. To that effect, data mining techniques may be useful, since they involve using tools to extract information among a large amount of data. This article presents the results of a study that aimed to develop and apply a visual method for systematic review based on concepts of Data Mining. This applied research, qualitative in its approaching and descriptive in its objectives, used a case study to demonstrate the usage of the proposed method. The results points to the efficiency of the method, having been possible to find and filter information in a systematic and productive way through it.

Keywords: design; data mining; methodology; scientific research

ISSN: 1808-3129

¹ Universidade Federal de Santa Catarina - UFSC - Brasil arinablum@gmail.com

² Eugenio Andrés Díaz Merino - UFSC/PPGEP/POSDESIGN - Brasil eugenio.merino@ufsc.br

³ UFSC/POSDESIGN - Brasil gisellemerino@gmail.com

Introdução

O lugar do design na ciência tem sido aprimorado com o uso mais efetivo de procedimentos técnicos e metodológicos reconhecidamente científicos. O rigor na busca da informação, na comunicação dela e, especialmente, no seu uso aplicado, torna-se fator relevante para comprovação de uma pesquisa enquanto dado científico.

A pesquisa científica, como afirmam Prodanov e Freitas (2013), deve ser sistemática, além de metódica e crítica, pois visa o conhecimento em torno de um ou mais aspectos de determinado assunto. Neste sentido, a revisão sistemática corrobora para reunir estudos semelhantes, aqueles de interesse do pesquisador conforme os estudos que está desenvolvendo, e sobre eles fundamentar a análise das informações e a ação crítica.

Neste contexto, relata-se neste artigo o estudo que teve como objetivo desenvolver e aplicar um método visual para revisão sistemática em design com base em conceitos da Mineração de Dados. Para tanto, como objetivos específicos, buscou-se estudar os conceitos da Mineração de Dados, organizá-los em forma de sinais visuais e realizar a aplicação em um estudo real.

Os objetivos surgiram a partir de uma questão de pesquisa que buscava um meio de como desenvolver visualmente um método que desse aporte ao processo de revisão sistemática em design. A questão, unida às informações sobre Mineração de Dados, levaram à hipótese de que o método seria uma eficiente ferramenta de apoio visual aos diversos momentos da investigação científica em que a revisão sistemática se faz necessária.

Numa abordagem qualitativa e descritiva, utilizou-se a aplicação em um estudo de caso, a fim de verificar a eficácia do método. O artigo apresenta os procedimentos metodológicos empregados, os conceitos da Mineração de Dados, o relato da caracterização do método proposto e conclui com a exemplificação a partir da aplicação em um caso real.

Procedimentos metodológicos

Para Flick (2009) a mudança tecnológica influenciou o próprio caráter da pesquisa qualitativa. Hoje, é possível fazer uso de diversos procedimentos que auxiliam no tratamento de uma informação qualitativa, entre eles os softwares especificamente direcionados para este fim. O uso crítico da informação, ou seja, o ato de converter um dado em conhecimento, não está centrado na tecnologia e, sim, no processo usado pelo pesquisador.

A pesquisa qualitativa, desta forma, pode se dar a partir do emprego de variados meios. Para esta aqui relatada, entre as apontadas por Flick (2009), fez-se uso das seguintes: codificação; armazenagem; busca e recuperação e mapeamento gráfico. Tais meios foram utilizados especialmente a fim de se construir a base do método.

A codificação envolveu a associação de palavras-chave que permitissem a recuperação futura dos dados. A armazenagem se deu no arquivamento dos dados

coletados, de forma organizada. A busca e recuperação relevou o princípio de manutenção e complementação do banco de dados, especialmente facilitando a sua localização para o caso da necessidade de, futuramente, os procedimentos serem repetidos. O mapeamento gráfico culminou no próprio método, com ênfase no fornecimento de elementos visuais auxiliares às buscas.

Pelo uso de tais meios e por envolver o teste do processo em um estudo de caso, a pesquisa se caracterizou como aplicada. Neste tipo de investigação, o objetivo é, conforme explicam Prodanov e Freitas (2013), a geração de conhecimentos que envolvem interesses para a solução de problemas específicos em uma aplicação prática.

O estudo de caso está sendo nomeado, nesta pesquisa, remetendo à aplicação prática. Nele, um estudo real, ou seja, o uso de dados de uma pesquisa científica em desenvolvimento, foi usado para testar o método. Todo o processo foi descrito na forma de passos constituintes do objeto de estudo – o método aqui apresentado.

Conceitos de Mineração de Dados

Uma revisão sistemática, segundo Flick (2013), segue critérios específicos e um foco estreito em torno de um tema geral de pesquisa. As direções da revisão sistemática são os próprios documentos de pesquisa. Sua utilização, explica o autor, implica em pré-definir critérios a fim de torná-la replicável e, em si mesma, passível de avaliação.

É, portanto, indispensável que, ao realizar uma coleta de dados, estes estejam primeiramente organizados e, preferencialmente, esquematizados de modo claro. A organização e a clareza darão suporte para que a pesquisa seja replicada – e assim confirmada ou contestada – e avaliada em termos de processo ou de qualidade.

Faz parte do processo de pesquisa científica uma revisão da literatura. Esta consiste em fazer uma verificação em torno do que já foi escrito e publicado sobre o assunto de interesse, quem escreveu e quando, se há lacunas e/ou oportunidades de abordagens que evitem a duplicidade de estudos no mesmo enfoque. (SILVA, 2005)

A revisão da literatura apoiada por um processo sistemático torna-se, então, relevante tanto para a localização de informações científicas quanto para a exploração consciente, não aleatória, dos dados. Ela determina o estado da arte e a corrobora para que os contornos do problema a ser estudado sejam definidos com mais precisão. (FLICK, 2013)

Por outro lado, a pergunta sobre “por onde começar?” faz parte desse processo. Ao se deparar com extensos bancos de dados e acesso facilitado às informações científicas disponibilizadas em variados meios, o pesquisador corre o risco de “se perder” no objetivo da pesquisa e, com isso, desfocar a busca. Com a não organização da busca, dados relevantes podem deixar de ser acessados ou, ainda, o pesquisador coleta um número tão grande de dados que não tem condições de analisá-los devidamente.

Neste sentido, a Mineração de Dados pode contribuir, munindo o pesquisador de passos e ferramentas que auxiliam na organização da busca e, com isso, em melhor uso da informação coletada. A Mineração de Dados ou Data Mining, conforme

Rezende (2005, p. 397), é o processo de extração de conhecimento a partir de grandes volumes de dados. Mais especificamente, explica a autora, visa “a obtenção de conhecimento útil e interessante, a partir de dados, para a utilização em um processo de tomada de decisão”.

Data Mining é uma das técnicas utilizadas para realização de KDD – Knowledge Discovery in Databases ou Descoberta de Conhecimento em Bases de dados – que é o processo de “garimpagem” de informações relevantes, que tem crescido e se complexizado com o aumento da capacidade de armazenamento em bancos de dados e o avanço das tecnologias de informação e comunicação. (QUONIAM et al., 2001)

Usando as técnicas de Mineração de Dados, diferentes tarefas podem ser executadas sobre as bases que guardam e compartilham informações científicas. Tais técnicas são úteis quando busca-se a geração de conhecimento emergido da associação por meio de aplicação de regras, da classificação dos dados coletados, da clusterização ou agrupamentos dos registros pela similaridade dos atributos ou características e, ainda, da visualização de anomalias com o objetivo de detectar desvios do padrão. (QUILICI-GONZALES; ZAMPIROLI, 2014)

No contexto que contempla o grande volume de dados armazenados – e hoje disponibilizados, de forma acessível, por instituições em sua maioria acadêmicas – e que embutem, muitos deles, informações realmente novas e interessantes, a Mineração de Dados se constituiu como forma interdisciplinar. Rezende (2005) aponta que Mineração de Dados tem estreitas relações com as contribuições oferecidas por diversas áreas, tais como Banco de Dados, Aprendizado de Máquina, Estatística, Recuperação de Informação, Computação Paralela e Distribuída.

Por esta interdisciplinaridade, a Mineração de Dados é explicada por diferentes perspectivas, que acordam com a sua área de abordagem. Citando autores de cada uma dessas áreas, Rezende (2005) explica que, sob a perspectiva da área de Banco de Dados, por exemplo, a Mineração de Dados trata do processo de descoberta de conhecimento armazenado em Base de Dados. De forma complementar, para a área de Aprendizado de Máquina, a Mineração de Dados é apontada como meio de extrair conhecimento potencialmente útil, implícito em dados e que era, no entanto, previamente desconhecido. Na perspectiva da área de Estatística, continua a autora, novas formas de dados podem emergir de resumos de relacionamentos inesperados de conjuntos de dados supervisionados em grandes quantidades.

Para Quoniam et al. (2001, p. 21), as ferramentas da Mineração de Dados identificam as correlações existentes nas fontes de dados e, neste sentido, através das técnicas de exploração de dados, extraem informações críticas. Estas, por sua vez, subsidiam o que os autores explanam como “processo decisório de uma organização”. Aplicado ao uso acadêmico e científico de uma informação extraída de bancos de dados relevantes do ponto de vista da pesquisa, tal processo decisório pode ser entendido como o uso do conhecimento na própria pesquisa, ou seja, na maneira de abordagem da ideia que acaba por dar subsídio a defesa de determinada linha de pensamento.

Em um artigo que apresentou o emprego de técnicas de Mineração de Dados para investigar os temas centrais da pesquisa em design em um periódico da área, Ribeiro et al. (2013) indicam que utilizaram softwares específicos para proceder a pes-

quisa. Esta se subdividiu em três principais momentos: a identificação do problema, o pré-processamento e a transformação.

O primeiro momento, identificação do problema, foi relacionado com o volume de informações, quando da grande quantidade de dados não se pôde, ainda, extrair conhecimento. Isso se deu porque o elevado número de informações torna inviável a análise por meio de métodos tradicionais, tais como gráficos e planilhas. Neste caso, a identificação de relações, quando existem, se limita às correlações.

No momento de pré-processamento, Ribeiro et al. (2013) usaram os procedimentos de organização dos dados coletados. No estudo por eles explanado, as ferramentas dos softwares foram integradas a partir do desenvolvimento de tabulações e planilhas. A “conversa” entre os softwares foi o primeiro passo para viabilizar a análise das informações.

Para tanto, os dados foram preparados no momento denominado de transformação. Foram decodificados ou transformados, de maneira que se viabilizasse o emprego de análises estatísticas ou de extração de padrões. Os agrupamentos e as identificações de correlações deste momento permitiram a definição dos resultados.

De fato, identificar problema, pré-processar e transformar dados são, basicamente, os passos que levam à Mineração de Dados. Rezende et al. (2005), neste contexto, explanam o processo de Mineração de Dados subdividindo-o em cinco etapas: (i) conhecimento do domínio; (ii) pré-processamento; (iii) extração de padrões; (iv) pós-processamento; (v) utilização do conhecimento.

O (i) conhecimento do domínio refere-se à seleção de dados em determinada base, considerando os objetivos da exploração. É a identificação do problema, quando devem ser respondidas questões importantes em relação às metas do processo, às restrições, aos critérios e aos objetivos com a extração do conhecimento. Esta etapa é fundamento para as seguintes.

O (ii) pré-processamento é o tratamento que os dados recebem a fim de serem utilizados nas etapas posteriores. Os dados são higienizados, ou seja, são tratados, limpos e sofrem redução de volume como meio de preparação para a extração de padrões. As transformações nos dados são efetivadas por meio da extração e integração, a organização e a limpeza, a seleção e a redução.

A (iii) extração de padrões tem o objetivo de encontrar modelos a partir dos dados. Visa a obtenção de resultados mais adequados aos objetivos a que se propõe o levantamento, ou seja, ao cumprimento das metas preestabelecidas. Nesta etapa, podem ser necessários ajustes que permitam maior precisão e melhor compreensão do conhecimento extraído. Para o devido ajuste de parâmetros, por vezes é necessário executar esta etapa mais de uma vez.

O (iv) pós-processamento é a avaliação quanto a qualidade dos dados selecionados. A passagem por esta etapa permitirá que os dados extraídos sejam considerados para a resolução de problemas reais. A fim de medir a qualidade dos dados – sua “compreensibilidade” e “interessabilidade”, como cita Rezende (2005, p. 412-413) – aspectos subjetivos podem ser filtros para a (v) utilização do conhecimento, que é aplicação prática do que foi minerado.

Método proposto

As cinco etapas apresentadas por Rezende et al. (2005) serviram de base para o método proposto. O primeiro passo para a construção visual do método foi estabelecer sinais gráficos que representassem as cinco etapas. Utilizando formas gráficas que fazem menção ao processo do trabalho de extração de minerais – e de onde surge o termo “Mineração” – e, ainda, a dinâmica indicada por Rezende et al, os passos do método foram visualmente organizados (Figura 1).



Figura 1 – Processo de Mineração de Dados: definição de sinais gráficos com base na organização apresentada por Rezende et al. Fonte: os autores.

As etapas do método foram, então, descritas de maneira aplicada. Em (i) conhecimento do domínio ficou estabelecido que o objetivo era selecionar as bases de dados que apresentavam um conjunto de informações relevantes, em termos de quantidade, frente ao interesse da pesquisa. As bases de dados podem variar de acordo com área de aplicação da pesquisa. No caso de uma revisão sistemática que envolva a área da Saúde, por exemplo, os periódicos focos do estudo são diferentes daqueles em que a pesquisa objetiva áreas como Arte ou Marketing. As bases de dados a serem escolhidas serão aquelas que, portanto, contemplam os periódicos de interesse, ou seja, diferentes para cada caso.

Como (ii) pré-processamento, os termos descritores foram definidos, bem como selecionados os limitadores de pesquisa disponibilizados em cada base de dados. Os descritores ou palavras-chave devem ser estabelecidos visando todo o contexto da pesquisa e considerando a sua escrita em língua inglesa. O cuidado em observar sinônimos e diferentes formas de abordagem de um determinado assunto deve ser parte desta etapa. É também neste momento que o estudo das bases de dados deve ocorrer a fim de se selecionar – nas ferramentas de busca disponibilizadas – os limitadores, tais como áreas de interesse da pesquisa, seleção de periódicos específicos, textos completos e/ou resumos críticos, entre outros.

Para (iii) extração de padrões, propôs-se que operadores booleanos fossem utilizados em combinação aos termos descritores, seguindo a lógica de interesse da pesquisa. Os operadores booleanos são usados junto aos termos descritores para relacioná-los e, assim, estabelecer as condições que se deseja nos resultados da pesquisa. O operador “AND” faz o papel de intersecção dos termos, “OR” de soma e “NOT” de exclusão (Figura 2). Alguns sistemas de busca disponibilizam os operadores para seleção por parte do usuário, outros requerem que o usuário os escreva no próprio campo de busca (ALEIXANDRE-BENAVENT et al., 2011).

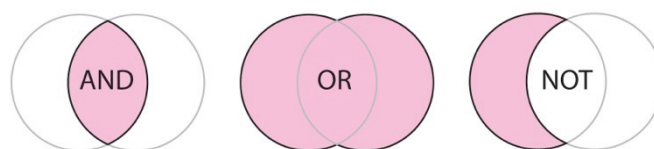


Figura 2 – Operadores booleanos AND, OR e NOT. Fonte: os autores.

O (iv) pós-processamento envolve a leitura e avaliação dos dados levantados e a seleção final que deve considerar, especialmente, a afinidade do dado com o contexto da pesquisa. Após a busca com base nos termos e operadores estabelecidos na etapa anterior, faz-se a leitura um pouco mais detalhada dos dados levantados – sem, no entanto, prender-se a uma análise em profundidade. O objetivo, neste momento, é relevar aspectos como título, palavras-chave e o resumo da publicação e, a partir disso, selecionar aqueles que são de interesse.

Como complemento na etapa (iv) pós-processamento e como preparo para a etapa (v) utilização do conhecimento, sugere-se que os dados selecionados sejam classificados e qualificados por grupos de relevância, de acordo com o foco da pesquisa. A qualificação dos grupos depende do interesse do pesquisador, por isso, ao classificá-los, sugere-se que os grupos recebam uma descrição.

Como exemplo dessa classificação, pode-se citar um grupo de dados selecionados a partir de uma busca de artigos em torno do tema “nanotecnologia em design de embalagem para alimentos”. Supondo que, hipoteticamente, 10 (dez) artigos neste assunto de interesse fossem localizados. Uma classificação, então, os separou em três diferentes grupos: 7 (sete) artigos tratavam de nanotecnologia em geral; 2 (dois) artigos relatavam a nanotecnologia em embalagens; 1 (um) artigo falava da nanotecnologia em embalagens de alimentos. Após descrever o que cada grupo contemplava, o pesquisador poderia qualificá-los da seguinte forma: os 7 (sete) primeiros em “tema geral”, outros 2 (dois) como “relevante” e 1 (um) como “foco principal”.

Com os dados coletados e classificados e/ou qualificados, é possível, então, passar à (v) utilização do conhecimento. Esta é a fase na qual procede-se à leitura, em profundidade, dos textos selecionados. A análise dos mesmos permitirá o seu uso tanto na construção da fundamentação teórica de algum estudo ou pesquisa quanto como meio de se chegar a outros dados não contemplados no procedimento de mineração que foi realizado.

Esses procedimentos perpassam, assim, pela transformação do dado – um elemento puro, quantificável – em informação, que é o dado analisado e contextualizado. Por sua vez, a informação passa a ser considerada conhecimento quando há

a compreensão, a análise e a síntese que levam ao seu uso inteligente (REZENDE, 2005). Neste sentido, ressalta-se que a etapa (v) utilização do conhecimento não é o “fim da pesquisa”, mas o meio para as próximas fases – já que essas dependem da revisão sistemática para que haja o uso inteligente das informações coletadas e, portanto, a viabilização do prosseguimento da pesquisa.

Aplicação do método

A fim de verificar a eficiência do método, o mesmo foi aplicado em um caso real. Numa pesquisa cujo objetivo era realizar uma revisão sistemática na área de gestão de design – especificamente abordando o tema do design gráfico para embalagens hospitalares de medicamentos – o método foi aplicado com o intuito de levantar referências em periódicos científicos.

Para tal levantamento, foram considerados, inicialmente, os parâmetros necessários para as três primeiras etapas – (i) conhecimento do domínio; (ii) pré-processamento; (iii) extração de padrões (Figura 3). O primeiro passo decorreu da escolha das bases de dados conforme descrição de suas abrangências e disponibilidades via Portal de Periódicos da Capes [3] com acesso através da Universidade Federal de Santa Catarina.

Assim, a pesquisa se aplicou na Ebsco Host [5], na Scielo [8], na Science Direct [6], na Scopus [7] e na Web of Science [19], por serem bases que contemplam um escopo multidisciplinar e englobam as grandes áreas de interesse do caso em estudo, especialmente Ciências Sociais Aplicadas (onde se enquadra o Design) e Saúde (frente ao foco em hospitais e medicamentos).

Passou-se, então, para a etapa (ii) pré-processamento, quando foi definido o período em que se pretendia resgatar os dados, o tipo de material a ser buscado, as áreas alvo da pesquisa e o estabelecimento dos descritores. Como período para resgate dos dados, buscaram-se as publicações efetivadas a partir do ano 2004.




Artigos em periódicos	
 (i) conhecimento do domínio	BASES DE DADOS: Ebsco Host Scielo Science Direct Scopus Web of Science
 (ii) pré-processamento	PERÍODO: a partir de 2004 TIPO: artigos completos em periódicos ÁREAS: Ciências Sociais Aplicadas, Saúde e afins DESCRITORES: design; gestão; gráfico; embalagem; medicamento; hospital.
 (iii) extração de padrões	COMBINAÇÃO DOS DESCRITORES: (design) AND (graphic) AND (management) AND (hospital) AND (medicine OR medication OR drug) AND (packaging OR package OR packing)

Figura 3 – Aplicação das primeiras etapas do método. Fonte: os autores.

Como tipo de material a ser buscado, determinou-se a pesquisa por artigos completos dentro das grandes áreas de interesse: Ciências Sociais Aplicadas, Saúde e afins – estas como principais. Algumas bases, contudo, permitiam o apontamento das áreas de interesse da pesquisa de forma mais específica e, de acordo com as abrangências nelas indicadas e a relação com o foco pretendido, foram assim selecionadas: Social Sciences & Humanities, Health Sciences e Life Sciences, na Scopus; Social Sciences, Arts & Humanities, Science Citation, na Web of Science; Engineering, Arts and Humanities, Business, Management, Accounting, Pharmacology, Toxicology and Pharmaceutical, na Science Direct.

Ainda na etapa (ii) pré-processamento, foram estabelecidos os termos descritores: design; gestão; gráfico; embalagem; medicamento; hospital (Figura 4). Posteriormente, foram organizados, de forma combinada, na etapa (iii) extração de padrões, quando se fez uso, para tanto, dos operadores booleanos "AND" e "OR". Os termos, em língua inglesa, foram combinados a fim de promover o levantamento de dados que abrangesse de macro para micro abordagem, conforme interesse da pesquisa.

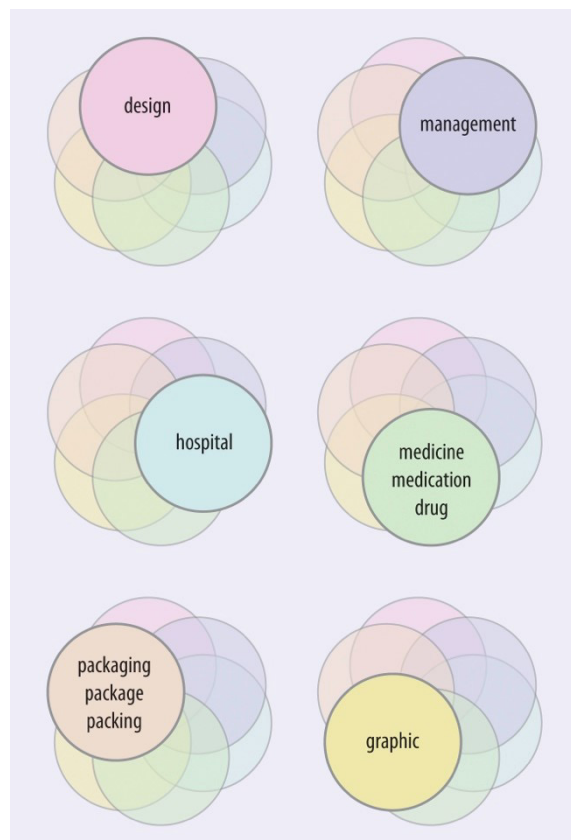


Figura 4 – Descritores definidos na etapa (ii) pré-processamento. Fonte: os autores.

O uso dos operadores booleanos permitiu uma abordagem organizada nos seguintes campos: (A) Gestão de design de embalagem; (B) Gestão de design de embalagem de medicamento; (C) Gestão de design gráfico de embalagem de medicamento; (D) Design de embalagem de medicamento; (E) Design gráfico de embalagem

de medicamento; (F) Design gráfico de embalagem de medicamento hospitalar; (G) Gestão de medicamento em hospital; (H) Gestão de embalagem de medicamento em hospital; (I) Gestão de design de embalagem de medicamento em hospital; (J) Gestão de design gráfico de embalagem de medicamento hospitalar.

Para uma visualização que contribuísse para o processo de busca, esses campos foram organizados com o uso de sinais e cores que funcionaram como um esquema de consulta no decorrer da pesquisa. A Figura 5 mostra a representação dos campos A, B e C; D, E e F; G, H e I; e J.

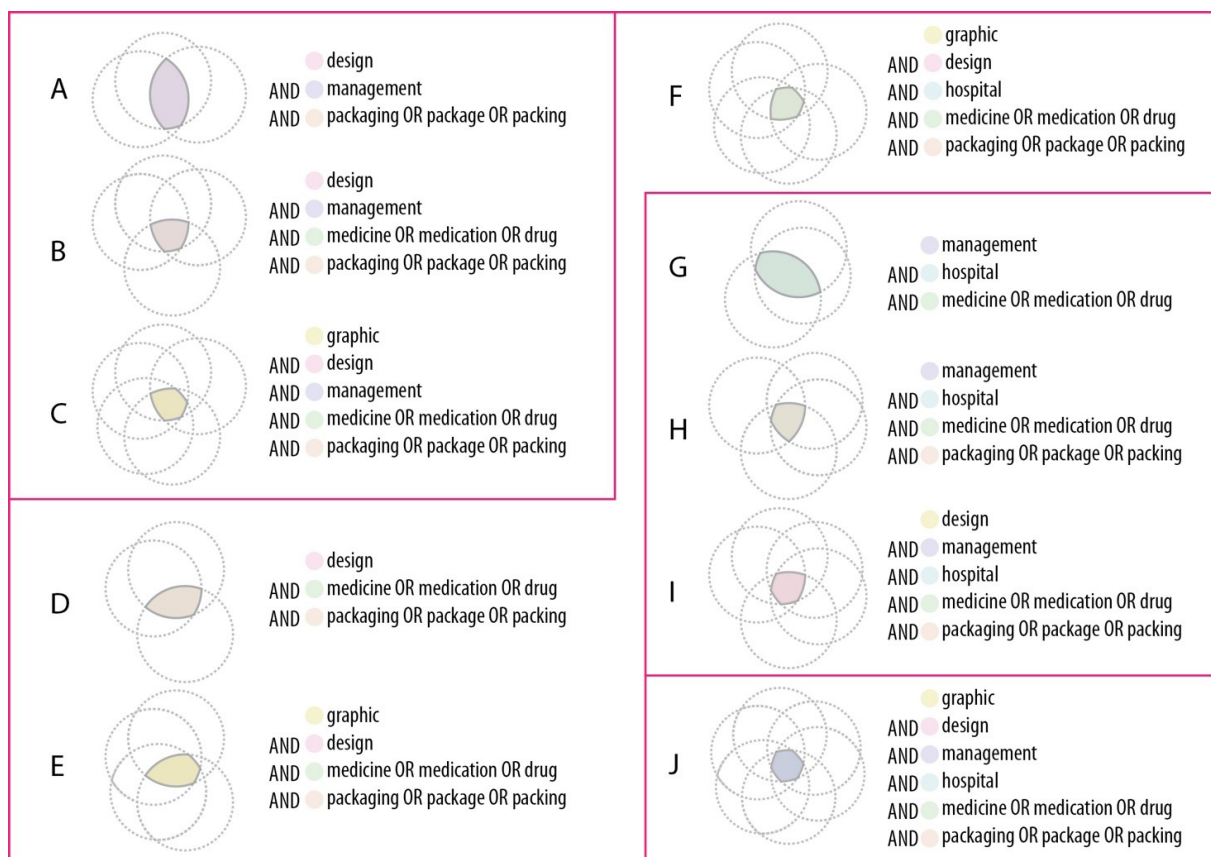


Figura 5 – Combinação de descritores na etapa (iii) extração de padrões. Fonte: os autores

Os campos A a C reuniram termos com foco na “gestão de design” e foram estruturados da seguinte maneira: A = [(design) AND (management) AND (packaging OR package OR packing)] ; B = [(A) AND (medicine OR medication OR drug)]; C = [(B) AND (graphic)]. Igualmente, os campos D a F – que reuniram termos com foco na “embalagem de medicamento” – seguiram a mesma lógica: D = [(design) AND (medicine OR medication OR drug) AND (packaging OR package OR packing)] ; E = [(D) AND (graphic)] ; F = [(E) AND (hospital)].

Idem às organizações dos campos anteriores, G a H reuniram termos com foco em “medicamento em hospital”: G = [(management) AND (hospital) AND (medicine OR medication OR drug)] ; H = [(G) AND (packaging OR package OR packing)] ; I = [H AND (design)]. O campo J englobou todos os descritores: J = [(graphic) AND (design) AND (management) AND (hospital) AND (medicine OR medication OR drug) AND (packaging OR package OR packing)].

A combinação e a divisão por campos permitiram a extração de padrões que re-

sultou em um levantamento quantitativo de artigos. A Figura 6 mostra os resultados do levantamento realizado entre os dias 01 de abril de 2015 e 03 de abril de 2015. Os números destacados correspondem aos artigos que foram analisados mais detalhadamente na etapa posterior.

Considerando a Mineração de Dados por meio dos agrupamentos, a etapa (iv) pós-processamento analisou mais detalhadamente os resultados numéricos levantados. Resultados acima de cem (indicados no quadro da Figura 6 por um asterisco) foram novamente minerados, a partir da análise de título e de palavras-chave. Posteriormente, os artigos foram apreciados através do resumo e enquadrados em “temas relacionados”. Destes, aqueles que tratavam de assuntos diretamente relacionados com a temática da pesquisa foram classificados como “relação de direta” (Figura 7).


Extração de padrões em periódicos científicos					
	Ebsco Host	Scielo	Science Direct	Scopus	Web of Science
A	2.842	14	23.827	17.512	2.608
B	428	0	4.391	6.012	146.935
C	2	0	409*	74	1
D	3.323	0	14.002	26.714	1.481
E	74	0	1.508	304	15
F	3	0	276*	49	0
G	146.935	714*	13.240	600.658	9.555
H	804	0	2.214	8.962	135
I	69	0	1.415*	2.455*	33
J	0	0	109*	39	0

Figura 6 – Quadro indicando o número de artigos levantados nas bases a partir da aplicação da etapa (iii) extração de padrões. Fonte: os autores.


Pós-processamento dos dados coletados						
	Ebsco Host	SciELO	Science Direct	Scopus	Web of Science	TOTAL:
Total etapa iii	74	728	2.209	2.568	49	
Nova mineração	5	1	6	48	6	
Temas relacionados	3*	1	3*	15*	3*	21
Relação direta	1*	0	2	1	1*	4
Semelhança de enfoque	0	0	0	0	0	0
*presença de artigos levantados em mais de uma base						

Figura 7 – Quadro indicando o número de artigos no assunto de interesse a partir da aplicação da etapa (iv) pós-processamento. Fonte: os autores.

Como resultado do processo aplicado, dos artigos indicados em “temas relacionados”, 4 (quatro) estavam disponíveis em mais de uma base. Os “temas relacionados” totalizaram 21 (vinte e um) artigos. Destes, 4 (quatro) apresentavam relação direta com o tema da pesquisa, ou seja, tratavam de estudos que envolvem elementos de design, de embalagem ou de rotulagem de medicamentos que se enquadravam em práticas aplicáveis a abordagem no contexto hospitalar ou mencionam esta: Chang, Lai e Chen (2012); Ward, Buckle e Clarkson (2010); Filik et al. (2006), Buckle et al. (2012). Estes artigos foram, então, separados para posterior uso na etapa seguinte, a (v) utilização do conhecimento, quando os textos coletados foram parte integrante de uma fundamentação teórica.

Considerações finais

Dado, informação e conhecimento são termos inseridos no processo de Mineração de Dados. O dado passa a ser informação científica quando coletado de maneira sistemática e quando, ao mesmo tempo, a partir da compreensão dos seus relacionamentos, estabelece-se uma compreensão sobre ele. O dado transformado em informação é útil a um processo de pesquisa porque ele se define como algo aplicável a determinado contexto e/ou a uma situação específica de estudo e investigação.

Por outro lado, a informação é ainda mais interessante, em termos científicos, no momento que ela passa a ser conhecimento. Dados se revertem de informação para conhecimento quando se dá o entendimento sobre eles, quando não só se compreendem os seus relacionamentos mas, mais do que isso, percebem-se e explicam-se os seus padrões.

Será a compreensão dos padrões a determinante para se chegar ao conhecimento em torno de um assunto em investigação. O entendimento dos princípios dos dados, das informações e do conhecimento é a maneira para embasar uma pesquisa, tanto a fim de não repetir o que outros pesquisadores já fundamentaram quando para, de fato, trazer novas contribuições à Ciência.

A passagem da coleta de dados para a informação e, em seguida, para o conhecimento científico é, portanto, um processo que exige sistematização a fim de que haja o devido rigor, do qual fazem parte a forma da coleta, o método e a crítica. Evidenciar tal sistematização no processo de pesquisa em design é importante para que se justifique a investigação e para que, criticamente, o Design continue a se estabelecer como área de conhecimento que preza pela técnica e pelos procedimentos.

O objetivo de desenvolver e aplicar um método visual para revisão sistemática em design com base em conceitos da Mineração de Dados foi pensado a fim de se estabelecer uma maneira de registro do processo de pesquisa, especialmente dos procedimentos de levantamento bibliográfico voltados ao Design. O objetivo foi alcançado com a proposição de um método visual simples, que se baseia no processo de Mineração de Dados, mas que não intenciona abrangê-lo de maneira complexa.

O estabelecimento do método a partir das cinco etapas do processo de Mineração de dados – (i) conhecimento do domínio; (ii) pré-processamento; (iii) extração de padrões; (iv) pós-processamento; (v) utilização do conhecimento – visou o fundamento conceitual desta forma de revisão sistemática, não cabendo a este estudo a intensão de abordar ferramentas específicas, tais como softwares de análise e outros instrumentos afins.

No caso relatado neste artigo, no qual o método foi aplicado, a eficiência ficou evidente especialmente em três pontos: na organização dos dados durante o processo de coleta e no momento de relatá-los; na facilidade de consultar ou ampliar os dados em caso de necessidade futura; na dinâmica perfeitamente aplicável a qualquer temática e a diferentes níveis de pesquisa.

A organização durante o processo de coleta auxilia para a agilidade no resgate dos dados, além de assegurar, para o pesquisador, que todos os passos foram devidamente seguidos. É também por meio da organização das informações que o rigor do processo se torna evidenciado e, com isso, o relato dos procedimentos metodológicos se apresenta com a devida veracidade.

Tendo o relato compatível com o processo realizado, é consequência que se dê a facilidade de consultar ou ampliar os dados em caso de necessidade futura. Num processo de pesquisa para uma tese, por exemplo, a revisão sistemática não acontece em um momento único, mas com periodicidade. Ter que voltar a consultar dados coletados em determinado processo, ter que refazer parte da revisão ou, ainda, ter que ampliar a busca por mais dados são parte da pesquisa e, por isso, o método organizado apoia tais ações.

Também, a dinâmica do método, aplicável a qualquer temática e a diferentes níveis de pesquisa, é relevante em um processo de investigação científica. A dinâmica permite que abordagens gerais sejam estudadas e, com a revisão sistemática, aprofundadas, detalhadas e especificadas. Além disso, o método não é restrito a casos específicos ou temas particulares dentro do design, mas é amplo e passível de ser

aplicado nos mais diferentes temas de pesquisa.

Os passos que englobam o método desenvolvido com base nos conceitos de Mineração de Dados – e apresentados neste artigo – mostraram-se eficientes no caso em que foi aplicado e aqui relatado. Além deste exemplo, outros temas foram também estudados utilizando-se este método, assim como fez-se o levantamento de dados em outras plataformas de pesquisa diferentes das aqui relatadas. Em todas as aplicações houve uso pleno do que foi proposto. Outros relatos futuros deverão apresentar tais casos.

Como método experimental, no entanto, entende-se que há limitação em relação ao distanciamento de ferramentas oferecidas no processo de Mineração de Dados e a maneira básica como estas foram abordadas no método proposto. Por outro lado, acredita-se que esta questão tende a ser melhorada no decorrer do uso deste método, já que novas ferramentas poderão ser agregadas, conforme a necessidade de aprofundamento da pesquisa. A continuação deste estudo está concentrada, portanto, na melhoria do método a partir da complementação com a inserção de outros instrumentos e de passos que contribuam, especialmente, para o uso das informações coletadas.

Referências

- ALEIXANDRE-BENAVENT, R.; ALCAIDE, G. G.; DIOS, J. G.; ALONSO-ARROYO, A. Fuentes de información bibliográfica (I) - *Fundamentos para la realización de búsquedas bibliográficas*. Acta Pediátrica Española, [S.L.], v. 69, n. 3, p.131-136, 2011.
- BUCKLE, P.; CLARKSON, P. J.; COLEMAN, R.; WARD, J.; ANDERSON, J. *Patient safety, systems design and ergonomics*. Applied Ergonomics, [S.L.], v. 37, n. 4, p.491-500, jul. 2006.
- CAPES. *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior do Ministério da Educação*. Portal de Periódicos da Capes. 2015b. Disponível em: <<http://www.periodicos.capes.gov.br/>>. Acesso em: 25 mar. 2015.
- CHANG, Ching-Hsiang; LAI, Yeong-Lin; CHEN, Chih-Cheng. *Implement the RFID Position Based System of automatic tablets packaging machine for patient safety*. Journal of Medical Systems, v. 36, n. 6, p.3463-3471, dez. 2012.
- EBSCO PUBLISHING. *Ebsco Host: Publicações científicas*. 2015. Disponível em: <<https://search.ebscohost.com/>>. Acesso em: 25 mar. 2015.
- ELSEVIER. *Science Direct*. 2015a. Disponível em: <www.sciencedirect.com>. Acesso em: 25 mar. 2015.
- _____. Scopus. 2015b. Disponível em: <<http://www.scopus.com/>>. Acesso em: 25 mar. 2015.

FAPESP et al. *Scielo: Scientific Eletronic Library Online*. 2015. Disponível em: <<http://www.scielo.org/>>. Acesso em: 25 mar. 2015.

FILIK, Ruth; PURDY, Kevin; GALE, Purdy; GERRET, David. *Labeling of medicines and patient safety: evaluating methods of reducing drug name confusion*. Human Factors, Santa Monica, Usa, v. 48, n. 1, p. 39-47, jun. 2006.

FLICK, Uew. *Introdução à metodologia de pesquisa: um guia para iniciantes*. Porto Alegre: Penso, 2013.

_____. *Introdução à pesquisa qualitativa*. Porto Alegre: Artmed, 2009.

SILVA, Edna Lúcia da. *Metodologia da pesquisa e elaboração de dissertação*. Florianópolis: Ufsc, 2005.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico*. Novo Hamburgo: Feevale, 2013.

QUILICI-GONZALES, José Artur; ZAMPIROLI, Francisco de Assis. *Sistemas inteligentes e mineração de dados*. São Paulo: Triunfal, 2014.

QUONIAM, Luc; TARAPANOFF, Kira; ARAÚJO JÚNIOR, Rogério Henrique de; ALVARES, Lillian. *Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil*. Ciência da Informação, Brasília, v. 30, n. 2, p.20-28, ago. 2001.

REZENDE, Solange Oliveira. *Mineração de dados*. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 25., 2005, São Leopoldo. Anais do XXV Congresso da Sociedade Brasileira de Computação. São Leopoldo: Unisinos, 2005. p. 397 - 433.

REZENDE, Solange Oliveira; PUGLIESI, Jaqueline Brigladori; MELANDA, Edson Augusto; PAULA, Marcos Ferreira de. *Mineração de dados*. In: REZENDE, Solange Oliveira. *Sistemas inteligentes: fundamentos e aplicações*. 2. ed. Barueri: Manole, 2005. Cap. 12. p. 307-336.

RIBEIRO, Vinicius G.; SILVEIRA, Sidnei R.; SILVEIRA, André da; ATKINSON, Rodrigo; ZABADAL, Jorge. *O emprego de técnicas de mineração de dados para definição de estratégias em processos de divulgação científica em periódicos de design*. Strategic Design Research Journal, Porto Alegre, v. 6, n. 2, p.85-94, ago. 2013.

THOMSON REUTERS. *Web of Science*. 2015. Disponível em: <<http://www.webof-science.com>>. Acesso em: 25 mar. 2015.

WARD, James; BUCKLE, Peter; CLARKSON, P. John. *Designing packaging to support the safe use of medicines at home*. Applied Ergonomics, -, v. 41, n. 5, p.682-694, set. 2010.