

O emprego de técnicas de mineração de dados para definição de estratégias em processos de divulgação científica em periódicos de design

The use of data mining techniques for defining strategies in scientific communication processes in design journals

Vinicius Gadis Ribeiro

vinicius@uniritter.edu.br

Centro Universitário Ritter dos Reis. Rua Orfanotrófio, 555, Alto Teresópolis, 90840-000, Porto Alegre, RS, Brasil

Escola Superior de Propaganda e Marketing. Rua Guilherme Shell, 350, São Geraldo, 90640-040, Porto Alegre, RS, Brasil

Sidnei Renato Silveira

sidneirenato.silveira@gmail.com

Universidade Federal de Santa Maria. Centro de Educação Superior Norte-RS. BR 386 km 40, Linha Sete de Setembro, 98400-000, Frederico Westphalen, RS, Brasil

André da Silveira

andre@um.pro.br

Centro Universitário Ritter dos Reis. Rua Orfanotrófio, 555, Alto Teresópolis, 90840-000, Porto Alegre, RS, Brasil

Rodrigo Atkinson

rodrigo.atkinson@accenture.com

Universidade de Santa Cruz do Sul. Av. Independência, 2293, Bairro Universitário, 96815-900, Santa Cruz do Sul, RS, Brasil

Jorge Zabadal

jorge.zabadal@ufrgs.br

Universidade Federal do Rio Grande do Sul. Escola de Engenharia, Departamento de Engenharia Nuclear. Av. Osvaldo Aranha, 99, 4º andar, Centro, 90460-900, Porto Alegre, RS, Brasil

Resumo

O objetivo deste trabalho é investigar os temas centrais da pesquisa em design por meio da análise de artigos do periódico Design Studies. Também visa identificar as tendências de evolução e futuro da pesquisa em design. Empregando análise bibliométrica e análise de rede, o trabalho analisa características de artigos do periódico Design Studies, buscando fatores que identificam os artigos de maior qualidade desse periódico. Devido à limitação da base de dados Scopus, este estudo considerou apenas artigos completos publicados em Design Studies no período de 2001-2011. A principal contribuição deste trabalho é a utilização de uma técnica de análise de rede para analisar métodos de pesquisa nos 317 artigos. Dentre os fatores identificados, destacam-se a possibilidade de parcerias, e temas voltados à Gestão de Design e Gráfica. De posse de informações sobre fatores de sucesso para publicação nesse periódico, torna-se possível destacar alguns procedimentos para submissão a esse periódico.

Palavras-chave: epistemologia do design, ciência do design, revisão sistemática, mineração de dados, análise de dados.

Abstract

The purpose of this paper is to investigate the core themes and research methods of design research by analyzing citations of papers in the journal Design Studies. It also aims to find out the evolution and future trends of design research. Employing bibliometrics and network analysis, the paper analyzes features of papers from Design Studies, looking for variables that identify factors of higher quality papers of this journal. Due to the limitation of the Scopus database, this study only managed to extract articles with full citations in Design Studies from 2001 to 2011. The main contribution of this paper is the use of a network analysis technique to analyze research methods in the 317 papers published in Design Studies between 2001 and 2010. Among the factors identified, we highlight the possibility of partnerships, and issues facing the Management and Graphic Web Design. Armed with information on success factors for publication in this journal, it becomes possible to highlight some procedures for submission to this journal.

Keywords: epistemology of design, design science, systematic review, data mining, data analysis.

Introdução

A comunicação científica é a fase de um processo de pesquisa em que o seu agente principal – o pesquisador – busca divulgar aos seus pares os resultados de seu trabalho. Frequentemente, essa divulgação formal se dá por intermédio do julgamento de seus pares em periódicos de boa qualidade ou eventos relevantes na área de investigação. Os trabalhos considerados de boa qualidade são aceitos para publicação ou para apresentação.

No Brasil, os Comitês de área da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – realizam reuniões periódicas para a definição dos periódicos relevantes de cada área do conhecimento que são relevantes, estratificando-os em oito níveis (Qualis A1 – mais elevado –, A2, B1, B2, B3, B4, B5 e C), baseando-se nas produções efetivamente publicadas pelos Programas de Pós-graduação *Stricto Sensu* de cada área do conhecimento. Nos estratos superiores, a CAPES define que a quantidade de periódicos A1 seja menor do que A2; que a soma desses se aproxime de 25%, e que a soma dos três estratos superiores seja no máximo 50% do total. Ademais, há quesitos específicos, em que cada área tem liberdade para definição, estabelecendo pontos de corte entre os extratos – referindo-se à periodicidade com que as edições são efetivamente publicadas; estruturação e diversidades institucional e geográfica do seu corpo de revisores; escrita em língua estrangeira – mais especificamente, inglês; quantidade de artigos por edição; diversidade de instituições de origem dos autores, buscando-se evitar periódicos com trabalhos endógenos – originados da mesma instituição de edição.

No caso dos cursos de Design – vinculados à área de Arquitetura e Urbanismo –, há diversos periódicos de mais alta qualificação, sendo o escolhido para o presente trabalho o *Design Studies*, da *Open University* (Londres). É assumido que esse periódico é o mais abrangente e interdisciplinar periódico da área de pesquisa em Design – com foco no projeto (Cross, 2010). Ademais, nessa área, é um dos mais antigos, e está disponível em formato digital desde 2001 – fato que viabilizou a realização do presente estudo. A política editorial do periódico para avaliação dos artigos possui duas fases: uma avaliação inicial realizada pelo editor, que avalia originalidade, fraquezas metodológicas, erros de língua inglesa ou problemas com escopo do trabalho em relação à revista. Após essa fase, artigos não rejeitados são submetidos a uma revisão do tipo duplo-cego. O tempo médio da primeira fase é de duas semanas, sendo o da segunda fase dois meses. Após a avaliação dos *referees*, em caso de aceitação, o autor recebe as recomendações dos avaliadores, sendo a decisão do editor com caráter definitivo. Essas políticas se encontram disponíveis na web (http://cdn.elsevier.com/promis_misc/JDST_Peer_Review_Policy.pdf).

Contudo, deve ser mencionado que a presença desse periódico no Qualis é uma situação de exceção: até o momento da coleta dos dados dos periódicos definidos no presente trabalho, os únicos dois artigos publicados por brasileiros se restringem a um trabalho apresentando histórico da primeira Escola de Desenho Industrial no Brasil e um recente trabalho de um brasileiro vinculado a uma instituição inglesa. Assim, observa-se que, até o momento

da coleta de dados, é raro um pesquisador brasileiro vinculado a uma Instituição de Ensino superior brasileira obter sucesso em ter seu trabalho aceito nesse veículo. Abstraindo-se a situação que, conforme a definição de periódico qualificado – deveria ter representativa publicação de brasileiros – é de se questionar quais são os fatores que potencializem o sucesso na aceitação de um artigo nesse periódico.

Dessa forma, o presente trabalho busca identificar possíveis fatores que potencializam a possibilidade de aceitação de um trabalho de investigação nesse periódico, empregando técnicas de mineração de dados. Como prova de conceito, foram empregados aplicativos de computador para facilitar o processo de análise de dados e de identificação de relações entre os campos de dados: mais especificamente, um programa de mineração de dados – Tanagra (Rakotomalala, 2005) – e um programa de interação de redes – Cytoscape (Cytoscape, 2013).

A estrutura do presente trabalho se encontra da seguinte forma: é apresentado breve referencial teórico sobre mineração de dados, além de breves referências aos métodos de pesquisa. Na Metodologia, encontra-se breve descrição de algumas ferramentas que podem ser aplicadas na mineração de dados, bem como a operacionalização do presente estudo. Nos resultados obtidos, detalha-se o estudo de caso realizado. Finalizando o trabalho, apresentam-se as considerações finais e as referências.

Referencial teórico

O processo de mineração de dados pode ser considerado como uma parte do KDD (*Knowledge Discovery in Databases* – Descoberta de Conhecimento em Banco de Dados), que foi definido por Fayyad como sendo um processo de constatação de padrões, que podem agregar utilidade ao conhecimento (Fayyad *et al.*, 1996).

Ele é composto por várias etapas interligadas, que vão desde a definição de domínio, seleção, preparação e transformação dos dados até a etapa de Mineração de Dados, em que se pode analisar os padrões descobertos e utilizar técnicas empregadas em Estatística e Banco de Dados para extração do conhecimento.

A primeira etapa do processo é a definição e a compreensão do domínio; após a definição, é necessário selecionar/criar os dados que serão trabalhados. Após o trabalho de seleção, trabalha-se com os dados que contêm representações ou informações inadequadas para o algoritmo a ser usado, excesso de atributos (redundantes ou desnecessários), atributos insuficientes, excesso de instâncias que podem afetar o tempo de processamento, instâncias insuficientes, instâncias incompletas (sem valores para alguns atributos). Essa etapa do Processo de Descoberta de Conhecimento é relevante para se conseguir fazer a mineração de dados – pois, como exemplo, alguns algoritmos de associações só trabalham com valores simbólicos/discretos.

Depois dos dados limpos, pré-processados, reduzidos – com o intuito de minimizar ruídos – e transformados de acordo com a proposta, define-se a técnica de Mineração de Dados a ser utilizada, assim como seus parâmetros, pois, dessa forma, os dados serão transformados de acordo com o algoritmo aplicado. Com o resultado disponível,

já se pode analisar, interpretar ou avaliar o conhecimento descoberto, podendo também repetir alguma etapa, se necessário.

Mineração de dados

"A Mineração de Dados é a análise dos conjuntos de dados observacionais, para encontrar relações insuspeitas e para resumir os dados de maneira compreensíveis e úteis para o proprietário dos mesmos" (Hand *et al.*, 2001, p. 1).

Neste contexto, a Mineração de dados – ou *Data Mining* – é um processo de extração de informações de uma grande base de dados para tomada de decisões e é aplicada em diversas áreas como empresas, pesquisas e indústrias, que utilizam os resultados, como, por exemplo, para melhoria de processos ou analisar tendências. Ele automatiza o processo de transformação e análise dos dados para descrever características do passado ou prever tendências do futuro. Possui ligação com outras técnicas e ciências, empregando diversos conceitos e técnicas de estatística, visualização, reconhecimento de padrões, processamento de alto desempenho, aprendizado por máquina, inteligência artificial, entre outras.

Para extrair o conhecimento, podem-se utilizar diversos métodos como Classificação, Modelos de Relacionamento entre Variáveis, Análise de Agrupamento, Sumarização, Modelo de Dependência, Regras de Associação e Análise de Séries Temporais, conforme definido por Fayyad *et al.* (1996).

Os métodos de Data Mining analisados neste trabalho foram:

(i) Classificação: é o aprendizado de uma função a ser usada para mapear dados em uma de várias classes discretas definidas previamente. Segundo Mattar (1998), a análise discriminante permite que dois ou mais grupos possam ser comparados com o objetivo de determinar se diferem uns dos outros e, também, a natureza da diferença, de forma que, com base em um conjunto de variáveis independentes, seja possível classificar indivíduos ou objetos em duas ou mais categorias mutuamente exclusivas. Dentre os métodos de Classificação podemos citar:

(a) Mínima Distância Euclidiana: A distância Euclidiana é a distância entre dois pontos, que pode ser provada pela aplicação do teorema de Pitágoras. No método, usa-se o protótipo de uma classe como assinatura e compararam-se atributos de uma instância com os protótipos. O protótipo mais próximo considerando a distância Euclidiana indica a classe.

(b) Árvores de Decisão: é uma estrutura de árvore, que faz representações simples do conhecimento, tem a função de particionar recursivamente um conjunto de treinamento até que cada subconjunto contenha casos de uma única classe. Os resultados obtidos, após a construção de uma árvore de decisão, são dados organizados de maneira compacta, com a árvore podendo ser utilizada para classificar novos casos.

Uma árvore contém nodos, que representam os atributos. Nessa estrutura, os arcos correspondem ao valor de um atributo e nodos folha, que designam uma classificação. A árvore pode ser lida a partir do teste encontrado em sua parte superior, normalmente chamado nó raiz da

árvore. Como exemplo de algoritmos que implementam a árvore de decisão, tem-se o ID3, o C4.5 e o C5.0.

(c) Vizinhos Mais Próximos: Nessa técnica, o conjunto de dados mais comum é mantido na memória para que os novos dados sejam comparados com estes. Se uma instância de classe desconhecida estiver a uma distância pequena de uma classe conhecida, é porque as classes devem ser as mesmas. Nesse método, não são criados protótipos ou assinaturas, usando-se as próprias instâncias.

(d) Redes Neurais: Uma Rede Neural Artificial (RNA) é uma técnica de construção de um modelo matemático, originalmente concebida com base no estudo do cérebro humano. Tem capacidade para aprendizado, generalização, associação e abstração. As redes neurais apresentam uma estrutura altamente interconectada e trabalhando em paralelo, assim como no cérebro humano. São compostas por processadores simples (neurônios artificiais) conectados entre si.

(ii) Regressão ou Predição: Aprendizado de uma função usada para mapear os valores associados aos dados. Observa-se, conforme Gujarati (2000), que o método dos mínimos quadrados ordinários tem propriedades estatísticas relevantes e apropriadas, que tornaram tal procedimento um dos mais poderosos e populares métodos de análise de regressão.

(iii) Agrupamento ou Clusterização: Identificação de grupos de dados onde os dados têm características semelhantes aos do mesmo grupo e onde os grupos tenham características diferentes entre si. Nesse tipo de análise, segundo Pereira (1999), o procedimento inicia com o cálculo das distâncias entre os objetos estudados dentro do espaço multiplano constituído por eixos de todas as medidas realizadas (variáveis), sendo, a seguir, os objetos agrupados conforme a proximidade entre eles. Na sequência, efetuam-se os agrupamentos por proximidade geométrica, o que permite o reconhecimento dos passos de agrupamento para a correta identificação de grupos dentro do universo dos objetos estudados. O algoritmo K-Médias pode ser destacado nesse método. Ele minimiza o erro quadrático calculado entre as instâncias e os centroides dos grupos.

(iv) Sumarização: Descrição compacta do que caracteriza um conjunto de dados (ex. conjunto de regras que descreve o comportamento e a relação entre os valores dos dados de meteorologia). As medidas de posição e variabilidade são exemplos simples de sumarização.

(v) Regras de Associação: Identificação de grupos de dados que apresentam concorrência entre si (ex. cesta de compras). A ideia é a derivação de correlações variadas que permitam subsidiar a tomada de decisões.

(vi) Séries temporais: Um conjunto de observações tomadas em tempos determinados, comumente em intervalos iguais (Murray, 1993). Entende-se que é uma sequência de observações sobre a variável de interesse, que é observada em pontos temporais discretos. A descrição do processo ou fenômeno se dá pela análise do comportamento.

(vii) Detecção de desvios ou *outliers*: Todas as técnicas de detecção de *outliers* fazem a seleção/identificação de dados que deveriam seguir um padrão ou comportamento esperado, mas não o fazem. Detecção de *outliers*, detecção de anomalias, detecção de ruído, detecção de desvio

e mineração de exceções são outras nomenclaturas para o método.

Métodos de pesquisa

O rigor metodológico de um trabalho vem da correta aplicação de um método de pesquisa. Essa aplicação se dá em todo o processo, desde a coleta de dados até sua análise. É uma das seções mais relevantes de um trabalho científico, visto que é a parte do trabalho onde é possível verificar a forma como o trabalho foi conduzido.

Breve descrição sobre os tipos de métodos de pesquisa, na área de Computação, é encontrada em Ribeiro e Zabadal (2010). Assim, buscou-se identificar com uma das variáveis de interesse qual foi o método de pesquisa empregado pelo pesquisador, do ponto de vista científico. Eram esperados, dentro dos trabalhos desenvolvidos na ótica positivista, estudos de caso (formalmente desenvolvidos), experimentos, *surveys* e estudos comparativos, entre outros. Estudos etnográficos, pela ótica interpretativa, também eram esperados. Em áreas que geram conhecimento a partir de suas práticas projetuais – como Design, Arquitetura, Computação e Engenharia – é costume trabalhar com a chamada dissertação-projeto, cuja participação no presente trabalho era esperada.

Metodologia

Há diversas ferramentas para mineração de dados disponíveis, em forma de *software* livre ou gratuitas – adequado para o projeto em questão. O módulo estatístico do Microsoft Excel, ROSETTA, Tanagra e WEKA são as ferramentas mais frequentemente empregadas pela comunidade acadêmica.

ROSETTA (ou *Rough Set Toolkit for Analysis of Data*) é um conjunto de ferramentas para análise de dados no âmbito da teoria dos conjuntos aproximativos. Foi desenvolvido pelo Knowledge Discover Group da NTNU (Norwegian University of Science and Technology), na Noruega, e o Logic Group, da Universidade de Varsóvia, Polônia. WEKA (ou *Waikato Environment for Knowledge Analysis*) é um ambiente que utiliza uma coleção de algoritmos de aprendizado de máquina e algumas tarefas de mineração de dados como: classificação, agrupamento e regras associativas e apresenta uma interface amigável para o usuário. Foi desenvolvido em Java, pelo Departamento de Ciência da Computação da Universidade de Waikato, na Nova Zelândia em 1999, constituindo-se como um *open source* (WEKA, 2013).

Já o Tanagra é um *software* livre de mineração de dados desenvolvido em Delphi por pesquisadores da Universidade de Lyon. É utilizado para fins acadêmicos e de pesquisa. Como é um *open source* (código aberto), os usuários podem acessar seus códigos e adicionar seus próprios algoritmos. Esse projeto é o sucessor do SIPINA, que implementou vários algoritmos de aprendizado supervisionado, especialmente uma construção interativa e visual de árvores de decisão. O Tanagra propõe diversos métodos de mineração de dados, análise exploratória e classificação estatística. Conta com processos de classificação supervisionada e não-supervisionada, tais como clusterização, análise fatorial, estatísticas parametrizadas e

não parametrizadas e regras de associação. Esse é um sistema integrado para análises estatísticas e de Mineração de Dados (Rakotomalala, 2005).

Para a atividade de Mineração de Dados, foi escolhido o *software* Tanagra, por dispor de uma coleção de algoritmos para diversas tarefas de mineração de dados, pelo seu uso ser livre e gratuito e pela facilidade de uso. Contudo, a identificação de relações básicas entre os dados torna-se uma atividade onerosa na Tanagra: a apresentação de dados referentes a relações se dá em valores numéricos, o que pode dificultar a visualização mais imediata e geral. Assim, para visualização específica de algumas informações básicas, foi empregado o programa *Cytoscape* – que, basicamente, é uma plataforma *open source* de bioinformática para visualização de redes de interação molecular e integração com perfis de expressão gênica e dados de outros estados. Especificamente, na situação descrita aqui, realizou a apresentação de dados de forma mais eficiente do que o Tanagra.

Operacionalização da metodologia

Para a preparação da base de dados, foi empregada uma planilha utilizando-se Microsoft Excel para a tabulação dos dados. A definição dos campos se deu em razão da identificação de cada um dos 317 artigos – tomada da base de dados Scopus ScienceDirect, entre os períodos de 2001 a 2011. A razão da definição desse intervalo de tempo foi a disponibilidade por parte da Instituição, além do fato de o periódico ter-se tornado disponível em formato digital a partir de 2001. A Figura 1 apresenta a planilha em questão – tanto a planilha onde os dados referentes a cada artigo publicado nesse intervalo foram tabulados, quanto a planilha higienizada – esta, ilustrada na Figura 2.

Identificação do problema

O volume de informações não permite que sua análise seja feita pelos métodos tradicionais (planilhas, gráficos, etc), já que com esses métodos podemos gerar relatórios, mas não a extração do conhecimento. Ademais, identificar relações entre os dados fica limitado apenas às correlações, quando existirem.

Pré-processamento

Como há integração entre as ferramentas Tanagra e Excel, os dados foram extraídos em formato .xls (formato nativo das planilhas do Microsoft Excel), sendo as tabelas já separadas por tabulações, o que permitiu ser importada pelo Tanagra sem alteração nenhuma na estrutura da planilha.

Transformação

Nessa fase de preparação dos dados, alguns campos podem ser recodificados ou transformados de forma a tornar viável seu emprego em análises estatísticas ou facilitar a extração de padrões.

A partir desse momento, efetuam-se os procedimentos de mineração de dados: identificação de colunas de

Figura 1. Planilha original, contendo a tabulação de dados no Excel.
Figure 1. Original worksheet containing the tabulation of data in Excel.

Figura 2. Dados estruturados – já higienizados.
Figure 2. Structured sanitized data.

dências – em especial com comportamento polinomial –, é possível notar o caráter periódico que alguns temas apresentam. Ainda considerando as linhas de tendências, pode-se observar que alguns temas se encontram em ascensão, tais como Gestão e Gráfica (*web*).

Uma situação observada é que a interface resultante do processo de análise do Programa Tanagra nem sempre se apresenta amigável, frequentemente apresentando os resultados de forma numérica. Dessa forma, a visualização geral fica dificultada. A Figura 4 apresenta uma interface desse tipo de consulta.

A partir dos dados quantitativos da ferramenta Tana-gra, passou-se a empregar a ferramenta Cytoscape. Essa ferramenta possibilitou a visualização de relações em rede de forma mais natural, a partir dos dados tabulados. A Figura 5 apresenta a relação país-técnica de pesquisa para o conjunto de dados analisado.

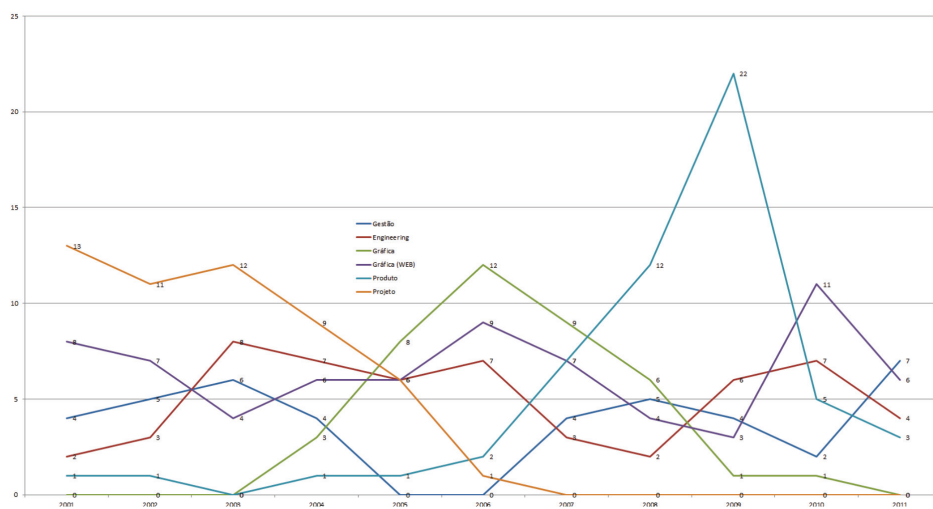


Figura 3. Quantidades de objetos de pesquisa por ano.
Figure 3. Amounts of research objects by year.

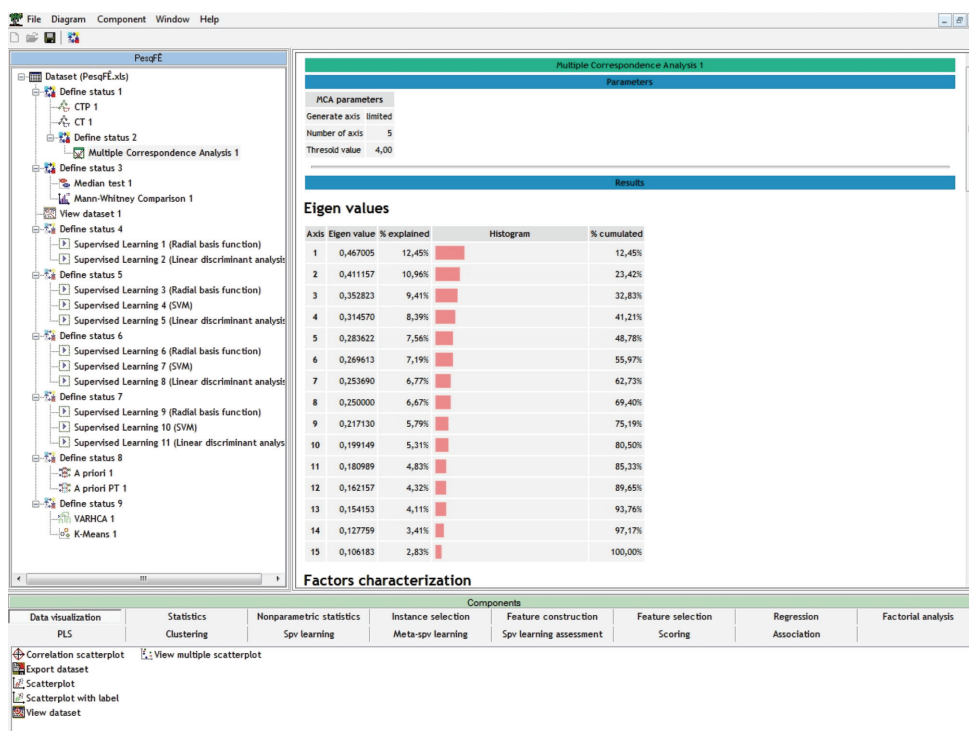


Figura 4. Apresentação dos dados do Tanagra.
Figure 4. Presentation of data from Tanagra.

Como se percebe, alguns países destacaram-se em grande quantidade de produções, assim como de técnicas de coleta de dados. Detalhamento sobre os dados da planilha apresenta que os países de língua nativa inglesa – Reino Unido (28,39%), Estados Unidos (18,93%) e Austrália (11,36%) – dominam as áreas, havendo pouco e específico espaço para algumas técnicas de pesquisa empregadas por outros países. Esse fato é lustrado na rede da Figura 6.

De forma similar à análise anterior com relação à visualização da relação entre os dados, procedeu-se com relação à possível relação entre os objetos de pesquisa e o país.

Na relação País-Objeto de pesquisa, quatro nações se destacam: Reino Unido, Estados Unidos, Holanda e Austrália. Essas nações produzem pesquisa com resultados relevantes investigando todos os objetos de pesquisa. *Projeto* é objeto de pesquisa desses principais países investigadores, além do Paquistão e da Finlândia. Por ser questão central no Design, era esperada maior quantidade de países investigando problemas nesse tema. Já nas pesquisas no objeto *Grafica (web)*, é visível vasta quantidade de países, incluindo os não-europeus – talvez uma interpretação seja a que não é uma pesquisa que impacte em custos para a

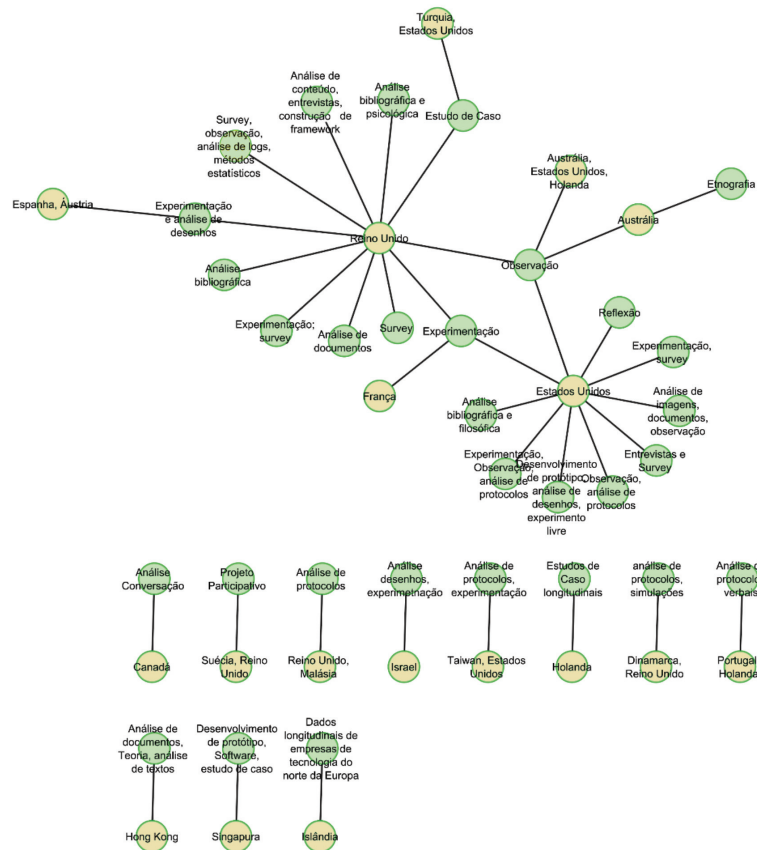


Figura 5. Representação gráfica da rede país – técnica de coleta de dados.
Figure 5. Network representation: country and technique of data collection.

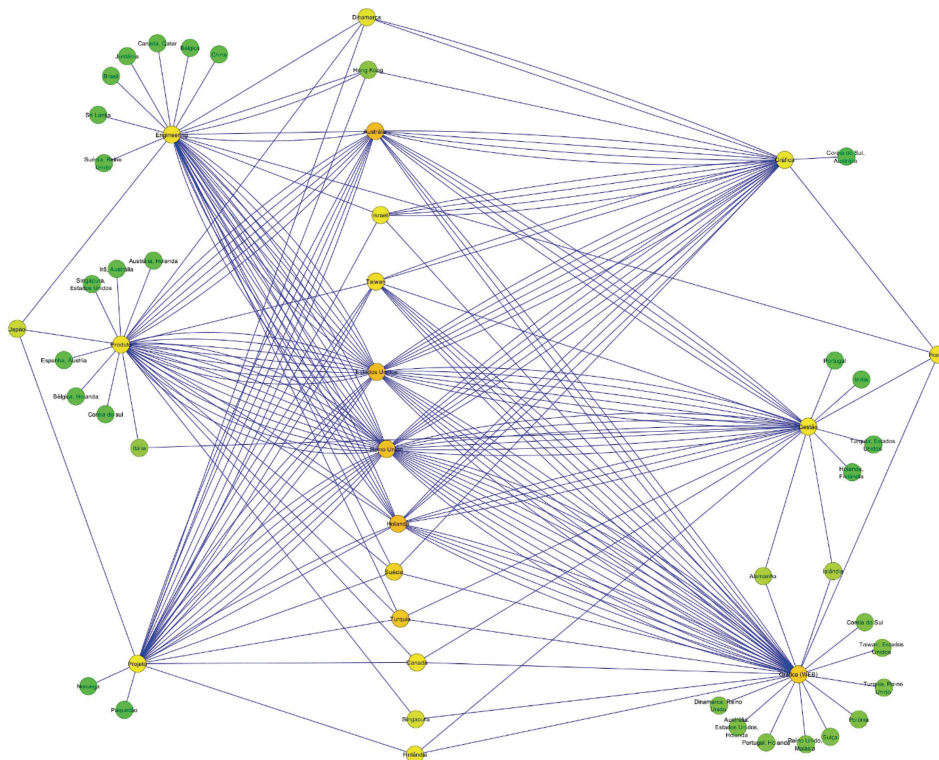


Figura 6. Representação gráfica da rede objeto de pesquisa/país.
Figure 6. Network representation: research object by country.

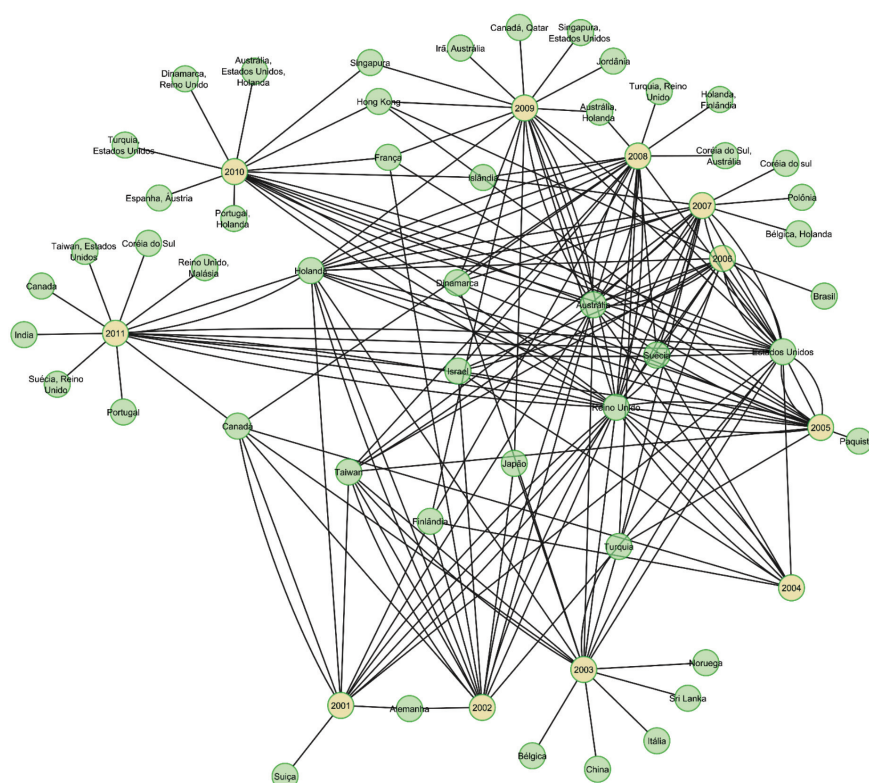


Figura 7. Representação gráfica de rede “Produção por países/ano”.

Figure 7. Network representation: “Country production by year”.

- Ano < 2006,5000
 - o Ano < 2004,5000
 - Ano < 2003,5000
 - Ano < 2001,5000 then Método = **Experimentação** (38,89 % of 18 examples)
 - Ano >= 2001,5000 then Método = **Reflexão teórica** (31,43 % of 35 examples)
 - Ano >= 2003,5000 then Método = **Estudo de Caso** (45,00 % of 20 examples)
 - o Ano >= 2004,5000
 - Ano < 2005,5000 then Método = **MultiMétodos** (36,36 % of 22 examples)
 - Ano >= 2005,5000 then Método = **Experimentação** (26,67 % of 15 examples)
- Ano >= 2006,5000 then Método = **MultiMétodos** (28,43 % of 102 examples)

Figura 8. Árvore de decisão obtida.

Figure 8. Decision tree obtained.

elaboração de protótipos. O objeto *Gestão* tem atraído pesquisadores de diversos continentes.

A Figura 7 reforça o que foi anteriormente verificado: alguns países são muito mais predominantes na produção acadêmica do periódico *Design Studies* – tais como Reino Unido, Estados Unidos e Holanda – através dos anos. Pode-se, pela figura, observar a quantidade de ligações em cada ano por país. Contudo, ao se observar a quantidade de artigos por ano na planilha de dados – ainda que detenha elevada produção – desses países, a contribuição norte-americana passou de 50 % a 60% e, atualmente, beira os 38%. Ao se observar os dados da planilha, verifica-se recentemente o incremento sutil, mas persistente, de países asiáticos e escandinavos.

Com relação à árvore de decisão, a relação identificada se deve ao ano de publicação – com o método de Experimentação e Estudo de Caso, além da abordagem multi-métodos obtendo maiores valores (Figura 8).

Considerações finais

Acredita-se que os objetivos propostos para o presente trabalho foram alcançados, tendo-se em vista que todas as etapas previstas foram realizadas: leitura crítica dos artigos dos periódicos, criação e população do banco de dados, estudo das ferramentas utilizadas, definição das técnicas de mineração a serem adotadas, pré-processamento dos dados, aplicação das técnicas de mineração

através da ferramenta Tanagra, aplicação de técnicas de visualização de informações através da ferramenta Cytoscape, análise dos dados minerados.

Algumas dificuldades foram encontradas, principalmente no que diz respeito à documentação da ferramenta Tanagra. Devido às dificuldades de se trabalhar com os métodos da ferramenta por limitações de documentação, foi dedicado muito tempo apenas para a escolha do tipo de mineração a ser realizado, uma vez que diversos testes e simulações foram realizados com a maioria dos algoritmos disponíveis na ferramenta até se encontrar um resultado satisfatório. Essa situação é comum nas ferramentas de mineração de dados – contudo, a ferramenta Tanagra apresenta esse inconveniente de não dispor de um sistema de ajuda, seja impresso, seja disponível no *software* ou na *web*. Já o trabalho com a ferramenta Cytoscape foi facilitado – de fato, a parte mais crítica foi a organização de informações visuais, visto que a ferramenta sempre devolve o resultado em uma forma matricial.

O emprego da ferramenta Tanagra mostrou-se mais complexo do que o esperado. Considerando-se que sua interface – e mesmo, o próprio programa – não contém sistema de ajuda, tornou-se mais difícil o seu uso. Apesar disso, algumas informações foram obtidas com o seu uso. Com relação ao método de pesquisa, destacam-se a experimentação (16,72%), estudo de caso (16,40%), e multimétodos (16,40%), seguidos de reflexão teórica (14,83%), dissertação-projeto (7,57%) e *survey* (5,36%). Deve ser destacado que, por abordagem multimétodos, os trabalhos mais frequentemente encontrados tratam de emprego de abordagens quantitativas para tomada de conhecimento de forma ampla no problema analisado e, posteriormente, em forma diferente de coleta e análise de dados, maior aprofundamento em situações particulares na amostra inicial.

Destaca-se que Gráfica (*web*) tem tido constante presença nas edições mais recentes dessa publicação, ao contrário de todas as outras. De certa forma, é curioso que o tema Projeto – tão relevante na área de Design – aparente ter queda de interesse/publicização. Esse fenômeno pode ser interpretado como a especialização do tema, afastando-se de investigações genéricas sobre Projeto, visto que o foco do periódico é o projeto. Já com relação ao objeto *Engenharia*, observa-se a publicação cuja origem, basicamente, são diversos países do terceiro mundo – o que pode sugerir que problemas desse tema sejam investigados nesses países.

Observa-se que, em algumas produções, há a relação de parcerias entre países – provavelmente, fomentada pelo período de formação em que um estudante estrangeiro realiza pesquisa em uma instituição desses países. Poucas produções de autores brasileiros têm sido encontradas no período avaliado (2001-2010), e um dos artigos de autor brasileiro em período posterior sugere essa participação.

Embora os métodos mais frequentes representem a produção na área – conforme se observa em outros periódicos e em eventos relevantes – esperava-se maior destaque no método da dissertação-projeto, sendo método característico por excelência nessa área. Embora considerada abordagem ainda empírica, esse método parte de uma realidade de um reconhecido problema da área e, empregando conhecimento – normalmente, de base

técnica –, gera-se a solução do problema, além de conhecimento específico da área.

A metodologia proposta de mineração de dados possibilitou encontrar de forma automática e quantitativa alguns padrões de conhecimento geral sobre os tipos de produção e suas relações e buscou encontrar padrões que possam ser úteis para a área. Considerando as informações obtidas, acadêmicos que desejarem comunicar os resultados de seus trabalhos nesse periódico deverão ater-se aos seguintes cuidados:

- Embora seja atenção elementar, a escrita no idioma inglês, de modo fluente, é fundamental;
- É relevante a busca por parcerias em pesquisa, o que é facilitado atualmente pelos programas governamentais com vistas à mobilidade acadêmica. A troca de dados, informações e conhecimento proporcionados por parcerias potencializa tanto a disseminação do conhecimento, quanto a análise de aspectos diversos – por exemplo, questões culturais em especificação de projeto – que podem trazer interessantes questões ao Design. Essa relação pode ser verificada pela análise do gráfico gerado pelo Cytoscape;
- Concentração dos esforços nas áreas de Gestão de Design e Gráfica (*web*), respeitando as linhas de pesquisa que cada pesquisador desenvolve;
- Reforço nas metodologias de Estudo de Caso e Experimentação. Essa constatação parte da análise individual dos artigos, visto que o rigor metodológico identificado nos trabalhos publicados é o esperado por um periódico desse nível de qualificação – ou seja, elevado. Não apenas dados estatísticos, mas análises de dados aprofundadas são comuns nos trabalhos observados.

Como trabalho futuro, a análise da mineração de textos – técnica que não foi trabalhada no presente estudo – dos resumos desses artigos pode revelar outras relações, mais detalhadas e de maior complexidade. Isso se deve, basicamente, ao rigor que o processo editorial desse periódico impõe, e que deve ser empregado no processo de qualquer periódico qualificado: em termos de texto, as partes componentes dos resumos de *Design Studies* pressupõem pelo menos contextualização, problema de pesquisa, método empregado e breve resultado da análise. É um processo relevante para a geração (ou descoberta) de conhecimento, corroborando com o que diversos autores afirmam – notadamente, Probst *et al.* (2002) e Takeuchi e Nonaka (2008). Contudo, para realizar esse tipo de análise, outra ferramenta de mineração de dados deverá ser empregada.

Referências

- CROSS, N. 2010. Editorial. *Design Studies*, **31**(1):1-3.
<http://dx.doi.org/10.1016/j.destud.2009.10.005>
- CYTOSCAPE 2002. Cytoscape. Disponível em: www.cytoscape.org. Acesso em: 5/10/2013.
- FAYYAD, U.M.; PIATETSKY-SHAPIO, G.; SMYTH, P. 1996. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 2, Portland, 1996. *Proceedings...* Portland, Oregon, **1**:33-47.

- GUJARATI, D.N. 2000. *Econometria Básica*. São Paulo, Makron Books, 812 p.
- HAND, D.; MANNILA, H.; SMYTH, P. 2001. *Principles of Data Mining*. Cambridge, The MIT Press, 425 p.
- KIERCZAK, M. 2009. A Rough Set toolkit for Analysis of Data. Disponível em: www.lcb.uu.se/tools/rosetta/. Acesso em: 11/09/2013.
- MATTAR, F.N. 1998. *Pesquisa de Marketing*. São Paulo, Atlas, 275 p.
- MURRAY, R.S. 1993. *Estatística*. São Paulo, Makron Books, 580 p.
- PEREIRA, J.C.R. 1999. *Análise de Dados Qualitativos*. São Paulo, Edusp/Fapesp, 160 p.
- PROBST, G.; RAUB, S.; ROMHARDT, K. 2002. *Gestão do Conhecimento: os elementos construtivos do sucesso*. Porto Alegre, Bookman, 288 p.
- RAKOTOMALALA, R. 2005. TANAGRA: a free software for research and academic purposes. In: ECOLE DE EXTRACCIÓN ET GESTION DES CONNAISSANCES, 3, Paris, 2005. *Proceedings...* Paris, RNTI-E-3, **2**:78-95.
- RIBEIRO, V.; ZABADAL, J. 2010. *Pesquisa em Computação: uma abordagem metodológica para trabalhos de conclusão de curso e projetos de iniciação científica*. Porto Alegre, Editora UniRitter, 203 p.
- TAKEUCHI, H.; NONAKA, I. 2008. *Gestão do conhecimento*. Porto Alegre, Bookman, 319 p.
- WEKA. 2013. Data Mining Software in Java. Disponível em: www.cs.waikato.ac.nz/ml/weka. Acesso em: 27/09/2013.

Submitted on August 20, 2013

Accepted on December 09, 2013