

Investigação em Educação em Ciência baseada em dados provenientes da internet

Francislê Neri de Souza

Universidade de Aveiro, Centro de Investigação em Didáctica e Tecnologia na Formação de Formadores
fns@ua.pt

Patrícia Albergaria Almeida

Universidade de Aveiro, Centro de Investigação em Didáctica e Tecnologia na Formação de Formadores
patriciaalmeida@ua.pt

Resumo

A rápida expansão da utilização da internet gerou novas experiências sociais que estão na origem de diferentes formas de interacção, de partilha de informação e de construção social do conhecimento. Todas estas experiências originam um volume incomensurável de dados de natureza diversa (vídeo, som, imagem, texto), susceptíveis de serem analisados sob diferentes perspectivas.

No entanto, como a própria internet, o campo emergente de estudos baseados na internet não está ainda estabelecido. As suas raízes são diversas e urge definir e caracterizar as metodologias e os métodos que mais se adequam a ambientes de investigação com características tão particulares e inovadoras. A natureza intangível da internet implica que a aplicação directa das metodologias tradicionais seja, normalmente, difícil e sem sucesso. Os investigadores vêem-se, assim, confrontados com novos modos de interacção que possibilitam o desenvolvimento de investigações com características diferentes das tradicionais e que, conseqüentemente, levantam questões metodológicas complexas.

O objectivo desta comunicação é analisar e discutir criticamente os tipos de investigações que podem ser desenvolvidas utilizando a internet, quer como meio de recolha de novos dados, quer como meio de colecta de dados disponíveis online, passíveis de serem analisados, e cuja análise pode contribuir para o aprofundamento da investigação em Educação em Ciência.

Palavras-chave: internet, tipos de investigação, tipos de análise qualitativa, corpus de dados, Educação em Ciência

Contextualização

Na sociedade actual, a utilização das tecnologias da informação e da comunicação, particularmente da internet, tem provocado alterações a níveis diversos, nomeadamente no que se refere às características espaciais e temporais das relações, gerando novas possibilidades de comunicação (Christians & Chen, 2004). O desenvolvimento destas novas formas de interacção tem influenciado o tipo de estudos desenvolvidos no âmbito da investigação em Ciências Sociais, nomeadamente em Educação em Ciência.

A utilização da internet como ferramenta de recolha de dados nunca foi tão frequente quanto hoje (Hewson, Yule, Laurent, & Vogel, 2003). Mesmo o tipo de

investigação associado a ambientes mais tradicionais, como é o caso dos estudos etnográficos, é actualmente passível de ser desenvolvido através da internet (Hine, 2004). Ward (1999) designou este tipo de estudos por *netnografia* e Stubbs (1998) por *cyber-etnografia*. Ward (1999) considera a netnografia como um meio de encarar, de forma séria, as interacções online e de as explorar através de um processo reflexivo que pode compreender um envolvimento activo em ambientes online. A mesma autora realizou uma investigação sobre comunidades online, estudando websites e fóruns interactivos, tendo adaptado os seus métodos às novas formas de comunicação com que se deparou (Ward, 1999).

São diversas as potencialidades e as novas oportunidades que as investigações mediadas pela internet (IMI)¹ parecem permitir. Algumas das vantagens associadas às investigações desenvolvidas em ou a partir de ambientes online, estão relacionadas com i) a facilidade com que os investigadores podem recrutar participantes para os seus estudos, podendo estabelecer amostras amplas e com características diversas; ii) a facilidade de comunicação (anónima, quando necessário) com os participantes; iii) elevados níveis de interactividade; iv) baixos custos envolvidos; v) a disponibilidade de meios para a gestão de cenários de investigação com níveis de complexidades diferentes (Christians & Chen, 2004; Hewson et al., 2003). No entanto, são vários os autores que defendem que muitos dos ambientes sociais que permitem o desenvolvimento de investigações online são resistentes às metodologias tradicionais. Algumas das dificuldades associadas a este tipo de investigação envolvem i) o nível reduzido de controlo por parte do investigador (sobre os participantes, os materiais e os procedimentos); ii) as limitações tecnológicas; iii) a necessidade de competências tecnológicas tanto por parte do investigador como dos participantes; iv) elaboração mais detalhada de ferramentas e procedimentos (Hewson et al., 2003).

Para muitos investigadores, a internet constitui ainda um ambiente pouco familiar e as suas condições espaciais e temporais requerem novos métodos e metodologias, que começam agora a ser desenvolvidos (Beddows, 2008), analisados e discutidos.

Investigação pela internet e com o apoio da internet

Hewson et al. (2003) distinguem *investigações primárias* de *investigações secundárias*, em contexto online. As primeiras fazem uso da internet para recrutar

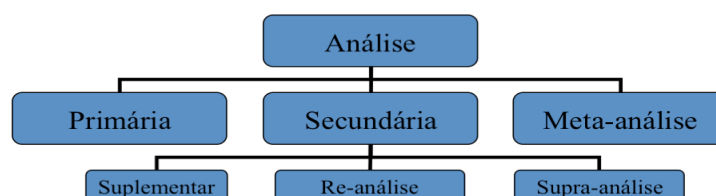
¹ Internet-mediated research (IMR)

participantes, recolher respostas, produzir e analisar novas evidências com base nos processos de comunicação. As investigações secundárias estão associadas ao uso da internet como meio de acesso a informações disponíveis online, tais como, as grandes bases de dados de revistas científicas e bibliotecas. Os mesmos autores enfatizam que “using the internet to locate secondary resources can have great pedagogical value and, we argue, enhance (but not replace) more traditional library based methods” (p.2).

Hewson et al. (2003), no livro *Internet Research Methods: a practical guide for the social and behavioural sciences*, analisam e reflectem sobre várias dimensões da investigação mediada pela internet, ou seja, sobre o uso da internet para promover interações e obter assim novos dados para dar resposta a uma determinada questão de investigação.

Também no que diz respeito ao tipo de análise efectuada é possível distinguir diferentes tipos, a saber: análise primária, análise secundária e meta-análise. Glass (1976) define *análise primária* como aquela que se refere à análise original dos dados, recolhidos especificamente para aquela investigação. Por sua vez, a *análise secundária* considera o uso de dados recolhidos no âmbito de investigações anteriores com o intuito de os reanalisar, diferenciando-se assim da *meta-análise*.

De acordo com Heaton (2004), a maior parte dos estudos que envolveram análise secundária de dados qualitativos são considerados estudos de *análise suplementar*, isto é, estudos que se debruçaram sobre aspectos que não foram considerados na análise inicial, ou que apenas foram considerados parcialmente. Muito menos comuns são *re-análises* das questões de investigação originais, que envolvem uma crítica à análise realizada inicialmente e as *supra-análises*, que se centram em novas questões de investigação.



Segundo Clark & Maynard (1998), a internet possui o potencial necessário para produzir um impacto significativo na natureza e no alcance da chamada análise secundária, mas actualmente os dados qualitativos disponibilizados online ainda são subvalorizados. A

análise secundária permite a reavaliação dos argumentos dos investigadores e permite também que investigadores que não tenham estado envolvidos no estudo original questionem e analisem os dados de modo diferente. Um dos principais motivos para realizar uma análise secundária de dados qualitativos prende-se com a riqueza deste tipo de dados, particularmente aqueles que são recolhidos em ambientes de investigação pouco comuns e de reprodução difícil ou não viável.

No entanto, o facto do investigador que conduz uma investigação secundária não ter participado na recolha inicial de dados (embora possa ter participado), impede-o de partilhar as mesmas perspectivas epistemológicas do investigador inicial, o que constitui uma barreira intransponível para a análise secundária (Mauthner, Parry & Milburn, 1998). Heaton (2004) prefere enfatizar o carácter metodológico eclético e pragmático da análise secundária, comparando-a a um trabalho de *bricolage*, uma vez que esta se baseia em múltiplos conjuntos de dados, fontes e abordagens metodológicas. Esta perspectiva atenua, assim, a importância atribuída ao envolvimento na investigação original e desafia a distinção entre análise primária e secundária.

Clark & Maynard (1998) incentivam a disponibilização online de dados qualitativos para que possam ser explorados por potenciais interessados. Já em 1978, Stenhouse sugeria um esforço colectivo no sentido do desenvolvimento de uma ‘história contemporânea’ da educação, na qual a qualidade da análise e a transparência da metodologia deveriam estar alinhadas. Deste modo, o mesmo autor propôs que os dados recolhidos deveriam ser explicitamente estruturados e apresentados, o que encorajaria a análise secundária. Deveriam também ser fornecidas informações sobre os métodos de recolha desses dados, de forma a permitir um envolvimento crítico com as estruturas e os processos interpretativos dos primeiros investigadores. Carmichael (2008) sugere ainda que além de disponibilizar os dados recolhidos e arquivados, os investigadores devem também disponibilizar os instrumentos, as notas de campo, e outro tipo de materiais, que permitem ter um conhecimento mais profundo sobre a natureza dos processos, das dificuldades e dos dilemas da investigação inicial.

Outra potencialidade da IMI, e particularmente da análise secundária, diz respeito ao papel que esta pode ter na formação de estudantes de pós-graduação e de futuros investigadores. Dados disponíveis na internet, e devidamente contextualizados, podem constituir uma oportunidade única para que futuros investigadores possam aprender a problematizar, a formular questões de investigação e a desenvolver análises científicas a

partir de dados reais e, conseqüentemente, complexos. É, assim, inquestionável o papel da análise secundária na aprendizagem sobre metodologias de investigação. Por outro lado, a IMI tem também um grande impacto nos processos de obtenção de dados bibliográficos para os estudantes de pós-graduação, dada a possibilidade de se obter um grande volume de dados num curto espaço de tempo e com custo reduzidos.

Verifica-se, assim, que a bibliografia relativa à investigação pela internet ou com apoio da internet se refere, fundamentalmente, a dois tipos de investigação:

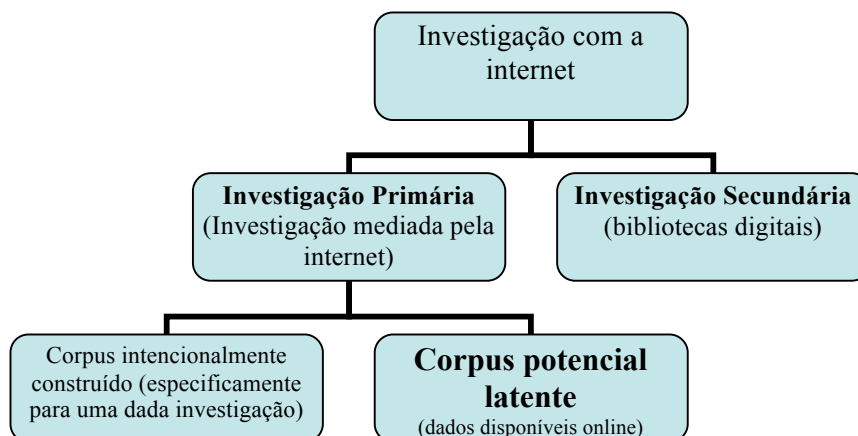
- i) um que se centra na recolha original de dados para dar resposta a uma determinada questão de investigação;
- ii) outro que se baseia em dados recolhidos anteriormente, e já analisados pelo menos uma vez; dados estes que podem ser reutilizados, constituindo assim o corpus de dados para uma nova investigação.

No entanto, e dado o novo tipo de interações que a internet facilita e medeia, parece-nos relevante considerar outro tipo de investigação, que tendo algumas características que se cruzam com as anteriormente referidas, possui outros aspectos distintivos, que colocam este tipo de investigação num patamar diferente. As metodologias de investigação, análise secundária incluída, têm evoluído no sentido de explorar e dar resposta aos desafios resultantes do crescimento e do desenvolvimento da internet. O estabelecimento dos maiores pontos de referência do panorama electrónico, tais como arquivos e bibliotecas, tem sido acompanhado pelo aparecimento de uma grande quantidade de redes sociais e recursos online. Arquivos pessoais, como por exemplo, blogs, websites, wikis, portfolios electrónicos, não só complementam os arquivos estabelecidos, mas também desafiam as distinções tradicionais entre o investigador que conduz o estudo original e o investigador secundário. A disponibilidade online deste tipo de recursos tem potencial para despoletar uma análise baseada em dados de fácil acesso, mas com características bastante diferentes das tradicionais. Assim, referimo-nos a um terceiro tipo de investigação que:

- iii) se centra em dados já existentes e disponíveis na internet, mas que não foram intencionalmente criados/recolhidos no âmbito de um projecto de investigação, isto é, não foram criados com o intuito de responder a uma questão de investigação, e que ainda não foram analisados. Referimo-nos, por exemplo, a dados disponíveis no YouTube, blogs, fóruns de discussão, sites para tirar dúvidas, wikis, etc.

Este é um tipo de investigação primária, mas que se distingue da investigação primária definida por Hewson et al. (2003), uma vez que, neste caso, os dados já existem. Referimo-nos a utilizar dados que se encontram disponíveis na internet, e que não foram produzidos com uma intencionalidade de investigação previamente estabelecida. Estamos interessados em dados da internet, que foram *produzidos de uma forma natural*, por utilizadores comuns da internet, e que designámos por *corpus de dados latente*. Propomos, então, a distinção entre investigação primária com *corpus de dados intencionalmente construído* e investigação primária com *corpus de dados latente*.

Embora a definição de corpus de dados seja mais complexa do que pode inicialmente aparentar, aqui consideramos como corpus de dados, uma colecção finita de materiais, determinada de antemão pelo investigador, e com a qual ele irá trabalhar (Barthes, 1967). De salientar, ainda, que o delineamento do corpus deve ser um processo cíclico, cuja construção está dependente da investigação empírica (Biber, 1993).



Também no que diz respeito à análise dos dados, este é um tipo de estudo que apesar de se entrecruzar com os tipos de análise tradicionalmente definidos e caracterizados, apresenta algumas características particulares. Não se trata somente de uma análise primária, uma vez que os dados já existem e se encontram disponíveis online. No entanto, também não se refere exclusivamente a uma análise secundária, uma vez que estes dados disponíveis na internet ainda não foram analisados.

A abundância de dados com potencial latente para investigação pode ser exemplificada partindo dos dados apresentados na Tabela 1.

Tabela 1 - Exemplos da abundância de dados disponíveis no YouTube (acesso no dia 20/05 de 2009)

Palavras-chave	Número de vídeos no Youtube	Número de comentários e avaliações do vídeo mais visto*	Número de exibições ao vídeo	Título do vídeo
Química	24800	(Com. 369) (Aval. 489)	470457	Bomba Atômica de Hiroshima e Nagasaki
Chemistry	36400	(Com. 444) (Aval. 332)	374941	Fun with Sodium Acetate
Biologia	14800	(Com. 32) (Aval. 78)	156690	Embriologia (Uma História de Amor)
Biology	31700	(Com. 149) (Aval. 195)	102220	Cell Biology
Física	57100	(Com. 171) (Aval. 222)	164941	Água que não cai / The water doesn't fall down
Physics	66700	(Com. 12725) (Aval. 18988)	5074657	Large Hadron Rap

* foi escolhido o primeiro vídeo mais visto que mostrava ter alguma relação com Ciência

Actualmente, começam a ser desenvolvidas, inclusivamente em contexto português (Fonseca & Gomes, 2007) algumas investigações baseadas em corpus com potencial latente, que facilmente ilustram o que referimos anteriormente. Apresentamos, de seguida, três exemplos de estudos, com características bastante diferentes, baseados em dados com potencial latente com origens distintas:

1) QuiProcura

De 1 de Março de 2005 a 28 de Fevereiro de 2007, foram recebidas 9654 mensagens de dúvidas no serviço de tutoria online sobre Química, QuiProcura (www.quiprocura.net). Deste total, Medeiros (2009) seleccionou 720 mensagens de acordo com o critério seguinte: selecção mensal de 60 mensagens de dúvidas, tendo sido realizada uma selecção aleatória. As mensagens foram analisadas quanto à data de envio e quanto ao conteúdo. Relativamente ao conteúdo foram considerados dois tipos de dados distintos: i) perfil dos usuários (género, idade, escolaridade e afiliação), ii) conteúdo das mensagens de dúvidas (solicitação, interrogação, imediatismo, contextualização e problemicidade) e conteúdo curricular (química geral, físico-química, química orgânica e outros assuntos).

A análise das mensagens de dúvidas revelou que a maior parte dos utilizadores eram estudantes que frequentavam ou haviam frequentado o Ensino Secundário, mas sem

vínculo ao Ensino Superior. O autor concluiu ainda que os usuários utilizavam o QuiProcura maioritariamente para resolverem trabalhos escolares.

2) YouTube

A 20 de Fevereiro de 2007, Keelan et al. (2007) fizeram uma pesquisa no YouTube (www.youtube.com) usando as palavras-chave *vaccination* e *imunization*, tendo seleccionado todos os vídeos em língua inglesa que contivessem qualquer tipo de mensagem relativa à imunização humana.

Consideraram 153 vídeos, e analisaram-nos quanto ao tipo, duração e argumentos científicos defendidos. Os mesmo autores contabilizaram o número de vezes que cada vídeo havia sido visto e a sua classificação (atribuída pelos utilizadores). Posteriormente, os vídeos foram categorizados como *positivos*, *negativos* ou *ambíguos*, *fundamentados* ou *não fundamentados*, de acordo com critérios definidos pelos autores. Os resultados obtidos remetem para a necessidade de os profissionais de saúde estarem preparados para responder a questões que possam resultar do visionamento dos vídeos estudados, e também para as potencialidades do uso do YouTube para promover uma comunicação efectiva entre os profissionais de saúde e a comunidade.

3) Ask-A-Scientist

Baram-Tsabari, Sethi, Bry & Yarden (2009) analisaram aproximadamente 79000 questões enviadas para o site Ask-A-Scientist (www.madsci.org), durante a última década. Foram considerados parâmetros como a idade, o género e o país de origem de quem colocou a questão, assim como o ano em que foi enviada. Os principais resultados mostraram uma dominância de contribuições femininas entre os alunos do ensino básico e secundário (embora esta dominância não se tenha verificado na amostra total), o que contraria o que se verifica em contextos offline, que se caracterizam por um maior interesse pela Ciência por parte do género masculino. Este entusiasmo feminino foi observado em diversos países, e não foi encontrada qualquer correlação entre o nível de igualdade dos géneros nesses países. Estes dados sugerem que a internet pode, potencialmente, desempenhar um papel democrático e uma forma de promover a igualdade, especialmente relevante em populações que, tradicionalmente, estão privadas de oportunidades iguais na aprendizagem formal de Ciência. Os autores deste estudo

verificaram ainda que à medida que os alunos avançavam na sua escolaridade, o interesse do género feminino em enviar questões para o site ia diminuindo relativamente ao interesse demonstrado pelos rapazes, confirmando, neste caso, o estereótipo característico de contextos presenciais.

Implicações para a Educação em Ciência

Tanto em contexto Português como em contexto internacional se tem verificado uma retracção relativamente à procura de formação superior na área de Ciências, e em certas áreas da Tecnologia. Considerando que o interesse é um dos mais poderosos factores de motivação em qualquer contexto de aprendizagem, reveste-se de especial importância investigar o foco de interesse dos estudantes quando abordam temas relacionados com ciência e tecnologia, e quais os factores que provocam o seu afastamento destas áreas de estudo.

A identificação e caracterização do foco de interesse dos estudantes pela Ciência pode ser feita através da investigação dos materiais produzidos por estes alunos e publicados na internet e, igualmente, através da investigação dos sites, blogs, wikis, vídeos no YouTube, mais visitados pelos estudantes, assim como as questões colocadas pelos mesmos em sites que pretendem esclarecer dúvidas sobre Ciência. A análise deste tipo de dados (corpus potencial latente) pode permitir também a identificação das principais dificuldades conceptuais e das concepções alternativas dos alunos. Parece-nos igualmente possível caracterizar os níveis de desenvolvimento cognitivo e emocional, assim como os estilos de aprendizagem dos alunos. Da mesma forma, os dados disponíveis na internet têm potencial para permitir a análise do nível de literacia científica, numa perspectiva de educação para a cidadania.

Referências bibliográficas

- Baram-Tsabari, A., Sethi, R. J., Bry, L., & Yarden, A. (2009). Asking scientists: a decade of questions analyzed by age, gender and country. *Science Education*, 93(1), 131-160.
- Barthes, R. (1967). *Elements of semiology*. London, Jonathan Cape.
- Bauer, M. W., & Aarts, B. (2000). A Construção do Corpus: Um Princípio para a Coleta de Dados Qualitativos. In M. W. Bauer & G. Gaskell (Eds.), *Pesquisa Qualitativa com Texto, Imagem e Som*. Petrópolis, RJ: Editora Vozes.
- Beddows, E. (2008). The Methodological Issues Associated with Internet-based Research, *International Journal of Emerging Technologies and Society*, 6(2), 124-139.
- Biber, D. (1993). Using register-diversified corpora for general language studies. *Computational Linguistics*, 19, 219-241.

- Carmichael, P. (2008). Secondary qualitative analysis using Internet resources. In N. Fielding, R. Lee, & G. Blank (Eds.), *The Sage handbook of online research methods*, 385-402. Thousand Oaks, CA: Sage.
- Clark, R. & Maynard, M. (1998). Research Methodology - Using Methodology for Secondary Analysis of Survey Research Data - "Act Globally, Think Locally", *Social Science Computer Review*, 16(1), 58-71.
- Christians, C. D. & Chen, S. S. (2004). Introduction: Technological environments and the evolution of social research methods, In M. D. Johns, S. S. Chen, & G. J. Hall (eds.) *Online Social Research: Methods, Issues & Ethics*, Peter Lang Publishing Inc. New York.
- Glass, G. V. (1976). Primary, secondary, and meta- analysis of research. *Educational Researcher*, 5(10), 3-8.
- Fonseca, L. F. C. & Gomes, M. J. (2007). Utilização dos blogues por docentes de Ciência: um estudo exploratório, In Barca, A., Peralbo, M., Porto, A., Duarte da Silva, B. e Almeida, L. (Eds.), *Libro de Actas do Congreso Internacional Galego-Portugués de Psicopedagogía*, Universidade da Coruña: Revista Galego-Portuguesa de Psicología e Educación.
- Hewson, C., Yule, P., Laurent, D., & Vogel, C. (2003). *Internet Research Methods: a practical guide for the social and behavioural sciences* (First ed.). London: SAGE Publications.
- Hine, C. (2004). Social Research Methods and the Internet: a thematic review, *Sociological Research Online*, 9(2).
- Keelan, J., Pavri-Garcia, V., Tomlinson, G. & Wilson, K. (2009). YouTube as a source of information on immunization: a content analysis, *JAMA*, 298(21).
- Mauthner, N., Parry, O. & Milburn, K. (1998). The data are out there, or are they? Implications for archiving qualitative data, *Sociology*, 32(4), 733-745.
- Medeiros, M. A. (2009). Análise de Mensagens Enviadas para um Sistema de Tutoria em Química na Web, *Química Nova na Escola*, 31(2).
- Stenhouse, L. (1978). Case study and case record: towards a contemporary history of education, *British Educational Research Journal*, 4(2), 21-39.
- Tubbs, S. (1998). Conflict and Co-Operation in the Virtual Community: eMail and the Wars of the Yugoslav Succession, *Sociological Research Online*, 3(3)
<http://www.socresonline.org.uk/3/3/7.html>
- Ward, K. J. (1999). The Cyber-Ethnographic (Re)Construction of Two Feminist Online Communities, *Sociological Research Online*, 4(1).
<http://www.socresonline.org.uk/4/1/ward.html>