

Getting Semantic: **Conforming our Journal to Times of Information Everywhere**

Yesterday, when I was a young scientist, life was simple, structured and easy. To learn what was new in my field, I used to go to a building called “the library”. Unless I had a specific goal, I would go through the most known *journals* of my area of interest or, better, through those I knew usually publish papers more interesting to me. Obviously, I did not search *all* journals of my specialty: there were too much of them, even in the library of my (small) university, which subscribed, say, only ten or fifteen percent of all which existed at the time.

Eventually, I would use a second resource of scientific information: the *indexers*, or the more complete and more interesting *abstracters*. There were only two or three of such resources in my field at that time. It was not too much a task to go through some years of scientific information during a busy and somewhat extended afternoon.

When looking for a specific article or for a narrow specialty, I frequently asked a librarian for help. Oh, these nice ladies (sorry, friends, I only remember lady-librarians...). I still love them for what they did for me. They had their (for me) mysterious ways to find anything! They always found the information I asked them for, even *thesis* and *dissertations* presented in far countries, oceans away...

Thesis and dissertations, by the way, was the closest I was able to come to ongoing research without leaving my home institution. The alternative was to attend a *national* or (may be once every two years...) an *international conference*. Ok, sometimes we had a *visiting researcher*, which delivered a talk. Mostly, however, the visitor discussed an already published paper, to “synchronize” the audience, and there was too few time left to effectively discuss the frontier of knowledge.

Oh, yes, one has not to forget what a library is historically about: *books*. They were all there, of all kinds, nicely organized in shelves. Review books, collecting information published in some hundreds of original papers in a “nervous” field. “Difficult” books at the graduate level. Classical, consolidated knowledge, in books written for undergraduate and introductory courses.

So, as I said at the beginning, my mind was well organized, in a simple and structured way. To each kind and level of information I needed, I had a specific source and a clear way to reach that source.

This nice time is now gone. I have no need to physically go to a place to search for scientific information. Yes, “the library” is still there, but stores only a very small amount of the information available in any field, and mainly the older part. For new, fresh information, I have the wonderful internet!

Wonderful, yes, but sometimes a nightmare! Although I still can find articles in *journals*, they may also be in an *open access repository*, in *virtual libraries*, in *portals* and even in *personal sites* - no longer nicely bound together, forming an issue with others on the same or on a neighboring subject. And the journals! A journal can be on its *own site*, or in a *portal* (like SciELO, www.scielo.org), or embedded in deep level directory of a *virtual library*... which, on its turn, also may contain single papers, thesis and dissertations, and books, and classroom material. However, there are also specific repositories for all this material, which can also be found in *personal or institutional sites*.... Information is everywhere, not any more nicely classified and separated by kind, subject, reliability, and “newness”.

Information is everywhere but, when I use a general search engine, it seems to me that being everywhere does not differ too much of being nowhere. Because my request returns 75635 potentially significant documents... of all kinds and types...in many languages... from all over the world... and I have to go through all them, many times judging by myself if the information is reliable. Oh, the good old times, when literature was referred and edited, and the filtering of quality was done by somebody else...

Internet, thus, helped a lot in accessibility and in quantity, but also raised two questions that have still to be answered: (i) how can I get just that what I need, and no more? and (ii) may I trust the information I get?

Let's postpone the second question to another time, and talk a little on what is happening concerning the first. The answer to it is the *semantic web*, which I like to oppose to the former *romantic web*. In the romantic times

of the web, building sites, creating electronic journals, exposing information in the internet, was like an adventure in the unknown, *boldly doing what no one has done before*, as Mr. Spock would say. But this model - or lack of - was unable to adequately comply to the needs of Science and of scientific information. The semantic web, and the concepts associated to it, came to "put some order in all this mess". Editors must know about it, but it is our feeling that, soon, all researchers will need to learn at least their basic concepts.

The main aim of the semantic web is to make the content of the internet more meaningful to machines and to humans. Many groups are involved in the construction of this new web. One important step was the world wide web consortium (www.w3c.com) development of XML (extensible markup language), which emphasizes content rather than appearance of web pages. A second and equally fundamental contribution comes from the Dublin Core metadata initiative (<http://dublincore.org>). Metadata are "information on information", and the Dublin Core is a set of fifteen basic (descriptive) metadata that should be associated to every information presented in the web. The third component is provided by the open archives initiative (www.openarchives.org) and its development of the protocol for metadata harvesting (OAI-PMH), through which information with appropriate metadata in web can be collected and meaningfully recovered from repositories built by data providers. There are still other relevant concepts, active research subjects in the field of information science, like ontology, the study and development of terminology that could be equivalently (in terms of the output of a search) entered in any language.

Although, as already said, we believe that every researcher will need, sooner or later, to become familiar with all this semantic web environment - we even think that information retrieval will be taught to all university freshmen in a short time - editors cannot ignore it already now! We, at the Interamerican Journal of Psychology, are working in two fronts towards the semantic web. The first one is on our new site, with a subject index, quick html abstracts of all papers since 1967, and not-so-fast but very useful pdf full text of papers, all with Dublin Core metadata (we will work backwards, starting 2005). The second is the adoption of the OJS (open journal system), developed at the University of British Columbia and translated into the Portuguese SEER (Sistema

Eletrônico de Edição de Revistas) by the Brazilian Institute of Scientific and Technologic Information (www.ibict.br). With this tool, the "compuscript" flux through the editorial process will be totally electronic, including referee analysis, editors comments, author corrections, and final version preparation. At the end, articles will be displayed in a web site in a OAI-PMH compatible way. This means to expose the knowledge developed by our authors within the highest standards of visibility available today, significantly enhancing the potential impact of Interamerican research in Psychology in academic and professional activities worldwide.

Piotr Trzesniak
Executive Editor

Sílvia Helena Koller
Editora

Tornar-se Semântico: ***Adequando nossa Revista ao Tempo de Informação por Toda Parte***

Há ainda poucos anos, quando eu era um jovem cientista, a vida era simples, estruturada e fácil de se viver. Para verificar o que havia de novo em meu campo, eu costumava ir até um prédio denominado “a biblioteca”. A menos que eu tivesse um propósito bem específico, eu percorria as *revistas* mais conhecidas de minha área de interesse ou, melhor, eu percorreria aquelas que eu já conhecia como habituais publicadoras de artigos que se revelavam mais interessantes para mim. Obviamente, eu não pesquisava *todas* as revistas da minha especialidade: elas eram demasiadas, mesmo na biblioteca da minha (pequena) universidade, que assinava, talvez, dez ou quinze por cento de todas as que existiam na época.

Eventualmente, eu recorrería a uma segunda fonte de informações científicas: os *indexadores*, ou os mais completos e interessantes *indexadores com resumos*. Destes, havia apenas dois ou três em meu campo naqueles dias. Não era excessivamente trabalhoso percorrer alguns anos de conhecimento científico em uma tarde dedicada, talvez apenas um pouco estendida.

Quando em busca de um artigo específico ou de um tópico altamente especializado, eu freqüentemente pedia auxílio a uma bibliotecária. Ah, estas agradáveis jovens (desculpem, amigos, eu apenas recordo bibliotecárias-mulheres...). Até hoje as amo pelo que fizeram por mim. Elas tinham lá os seus caminhos (para mim) misteriosos para encontrar qualquer obra! Elas sempre localizaram a informação por mim solicitada, mesmos *teses* e *dissertações* defendidas em pequenos e distantes países.

Teses e dissertações, a propósito, era o mais próximo que eu era capaz de chegar da pesquisa em andamento sem deixar a instituição em que trabalhava. A alternativa eram os *encontros nacionais* e, talvez uma vez a cada dois anos, uma *conferência internacional*. Sim, de quando em vez tínhamos um pesquisador visitante, que apresentava um seminário. Quase sempre, no entanto, o visitante discutia um artigo já publicado, para “sincronizar” a audiência, e sobrava muito pouco tempo para efetivamente discutir a vanguarda da ciência.

Sim, claro, não se pode esquecer a finalidade histórica das bibliotecas: *livros*. Estavam todos lá, de todos os tipos, lindamente organizados nas prateleiras. Livros de revisão, capturando e sistematizando a informação publicada em centenas de artigos em uma área “nervosa” do conhecimento. Livros “difíceis”, escritos para a pós-

graduação. O conhecimento clássico, consolidado, preparado para os níveis introdutório e avançado da graduação.

Assim como eu disse no princípio, minha mente estava bem organizada, de um modo simples e estruturado. Para cada espécie e nível de informação de que eu precisava, eu tinha uma fonte bem definida e um caminho bem conhecido que até ela iria me conduzir.

Estes bons tempos agora não existem mais. Eu não mais preciso me deslocar fisicamente para um determinado local em busca de informação científica. Sim, “a biblioteca” ainda existe, mas guarda apenas uma pequena parte da informação disponível, qualquer que seja o campo, e principalmente a parcela mais antiga. Para a informação mais nova, recente, eu tenho a maravilhosa *internet!*

Maravilhosa, sim, mas muitas vezes um pesadelo! Embora eu ainda possa encontrar artigos em *revistas*, eles também podem estar em *repositórios de acesso livre*, em *portais* e até em *páginas pessoais* – não mais belamente encadernados, formando um fascículo com outros sobre o mesmo assunto ou um tema afim. E as próprias revistas! Uma revista pode estar em seu *próprio site*, ou pertencer a um *portal* (como a SciELO, www.scielo.org), ou encontrar-se num diretório de nível profundo de uma *biblioteca virtual*...a qual, por sua vez, pode também conter artigos isolados, teses, dissertações e material didático, o qual, porém, pode ser também encontrado em *sites pessoais ou institucionais*... A informação está em toda parte, não mais tranqüilamente separada e organizada por espécie, assunto, confiabilidade e atualidade.

A informação está em toda parte, porém, quando eu uso um mecanismo geral de busca, parece-me que estar em toda parte não difere muito de estar-em-lugar-qualquer. Pois meu pedido devolve 75635 documentos potencialmente de interesse... de todos os tipos e espécies... em vários idiomas... de todas as partes do planeta... e eu terei de percorrê-los todos, muitas vezes tendo que decidir eu mesmo se aquela informação é confiável. Ah, os bons e velhos tempos, quando a literatura científica era arbitrada e editada, e a filtragem da qualidade era feita por outras pessoas, antes que ela chegasse até mim...

A *internet*, desse modo, ajudou muito quanto à acessibilidade e à quantidade, mas também trouxe duas questões que ainda precisam ser mais bem respondidas: (i)

como eu poderei recuperar exatamente e apenas aquilo de que preciso? e (ii) posso confiar na informação recuperada?

Vamos deixar a segunda questão para um outro momento, e comentar um pouco o que está sendo feito relativamente à primeira. A resposta a ela é a *web semântica*, que eu gosto de contrapor à já ultrapassada *web romântica*. Nos tempos românticos da *web*, construir *sites*, criar revistas eletrônicas, expor informação na *internet*, era como uma aventura rumo ao desconhecido, *audaciosamente indo onde nenhum ser humano jamais esteve*, como diria o Sr. Spock. Porém este modelo – ou esta falta de modelo – era incapaz de atender adequadamente às necessidades da Ciência e da informação científica. A *web semântica* e os conceitos a ela associados vieram colocar alguma ordem nesse caos. Editores tem que tomar conhecimento disso, mas é nossa impressão que, em breve, todo pesquisador terá que familiarizar-se ao menos com os seus fundamentos.

O grande objetivo da *web semântica* é tornar o conteúdo da *internet* mais claro e significativo simultaneamente para as pessoas e para as máquinas. Diversos grupos estão envolvidos na construção dessa nova *internet*. Um importante passo foi o desenvolvimento da XML (*extensible markup language*), que enfatiza não mais a aparência, porém o conteúdo das páginas *web*. Uma segunda e igualmente fundamental contribuição vem da Dublin Core *metadata initiative* (<http://dublincore.org>). *Metadados são informações sobre a informação*, e o Dublin Core é um conjunto de quinze metadados (descritivos) básicos que devem ser associados a toda informação disponibilizada na *internet*. A terceira componente é devida ao movimento dos arquivos abertos (*open archives initiative*, OAI: www.openarchives.org): consiste no protocolo para coleta de metadados (*protocol for metadata harvesting*, PMH), através do qual toda a informação na *web*, que esteja adequadamente descrita por metadados, pode ser coletada e apropriadamente recuperada de repositórios construídos por provedores de dados. Existem ainda outros conceitos relevantes, que correspondem a campos de pesquisa ativos na área de ciência da informação. Por exemplo, o estudo de ontologias visando, entre outros objetivos, ao desenvolvimento de terminologias que se equivalham (em termos da devolução de documentos em uma pesquisa na *web*), independentemente do idioma em que sejam originalmente introduzidas.

Embora acreditemos, como já dito, que todo pesquisador precisará, cedo ou tarde, familiarizar-se com todo este ambiente semântico da *web* – nós

acreditamos, inclusive, que a recuperação de informações brevemente deverá ser uma disciplina básica em cursos universitários de todas as especialidades – os editores não podem ignorá-lo desde já. Nós, do *Interamerican Journal of Psychology*, estamos trabalhando em duas frentes rumos à inserção na *web semântica*. A primeira é o nosso novo *site*, com um índice de assuntos, resumos de todos os artigos desde 1967 em (rápido!) html e textos integrais em nem-tão-rápido porém muito útil pdf, tudo com metadados no padrão Dublin Core (trabalharemos recuando a partir de 2005). A segunda é adoção do *open journal system*, desenvolvido pela Universidade de British Columbia, no Canadá, e traduzido para o português pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (www.ibict.br), sob a denominação Sistema Eletrônico de Editoração de Revistas. Com o emprego desta ferramenta, o fluxo dos “compuscritos” através do processo editorial será totalmente eletrônico, incluindo o processo de revisão pelos pares, análise pelos editores, correção pelos autores e preparação da versão final. Através dela, também, os artigos serão disponibilizados em um *site* com metadados compatíveis com o protocolo OAI-PMH. Isto implica expor o conhecimento desenvolvido pelos nossos autores dentro dos mais elevados padrões de visibilidade atualmente disponíveis, significativamente ampliando o potencial de impacto da pesquisa interamericana sobre a Psicologia acadêmica e profissional em nível mundial.

Piotr Trzesniak
Editor Executivo

Sílvia Helena Koller
Editora