

VISUALIZANDO BASES CURRICULARES

DA SILVA, Luciano Carlos Ribeiro¹; **DO NASCIMENTO**, Hugo Alexandre Dantas²
Instituto de Informática

¹luciano@inf.ufg.br; ²hadn@inf.ufg.br

Palavras chave: *cienciometria; visualização de informações; Lattes; bases curriculares.*

1. INTRODUÇÃO

Anualmente, centenas de trabalhos em C&T (Ciência e Tecnologia) são concluídos pelas várias instituições de pesquisa no Brasil e no Mundo. Obter um panorama dessa produção é importante tanto para governos quanto para as próprias instituições e pesquisadores envolvidos, uma vez que tais informações ajudam a definir metas e políticas de trabalho para o desenvolvimento científico.

O Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) mantém um conjunto de sistemas de informações, bases de dados e portais Web voltados para a gestão de C&T (Ciência e Tecnologia), chamado de Plataforma Lattes. Nessa plataforma, acadêmicos e pesquisadores podem gravar seus currículos, relacionando toda a sua produção técnico-científica. O objetivo do presente trabalho é propor um modo de se ampliar o aproveitamento das informações contidas na plataforma Lattes, através de uma maximização da expressividade de seus dados.

O restante deste documento está organizado como segue: a Seção 2 traz uma revisão sucinta dos principais temas relacionado ao trabalho. A Seção 3 descreve um protótipo criado para a exploração visual de bases curriculares. Finalmente, as conclusões do trabalho são comentadas na Seção 4.

2. METODOLOGIA

2.1. FUNDAMENTAÇÃO TEÓRICA

2.1.1. Cienciometria

A tarefa de se criar indicadores coerentes para mensurar a produção científico-tecnológica não é trivial, em decorrência da subjetividade do seu objeto de trabalho. Não se mede facilmente o fruto do intelecto, ao contrário dos demais ramos de atividades sociais, como o comércio e a indústria, os quais dispõem de diversos indicadores absolutos para quantificar e comparar produtividade. A simples contagem dos textos produzidos por pesquisadores, embora amplamente adotado pela **bibliometria**, deixa de lado inúmeros fatores relevantes que, se não forem contabilizados, podem mascarar problemas e até mesmo induzir a graves erros de análise. É nesse contexto de incertezas que surge a **Cienciometria**, a qual busca na estatística, sociologia e filosofia técnicas sofisticadas para medir a ciência. Pode-se definir a cienciometria como sendo a ciência de se medir a ciência.

Podemos classificar os indicadores de produção adotados na cienciometria como: **income** - que mede os investimentos financeiros em C&T; **outcome** - que mede os resultados obtidos; **indicadores-produto** - que medem de forma imediata o grau de produção, mas não levam em conta a sua relevância;

indicadores-impacto - que mede o grau de relevância da produção, o que só é possível a longo prazo.

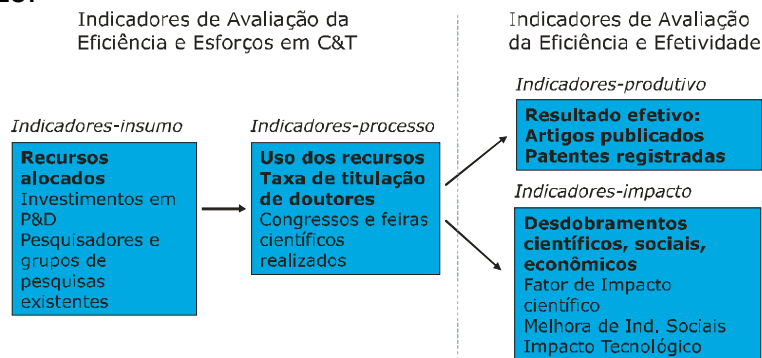


Figura 1. Diagrama das classes dos indicadores cienciométricos.

Os principais indicadores da ciencimetria são: **números de trabalhos** - não leva em conta a sua relevância; **número de citações** - muito utilizado pela ciencimetria, pois consegue indicar o impacto de um trabalho na sua comunidade; **co-autoria** - identificar como os atores cooperam; **números de patentes** - inovações tecnológicas; **relevância do meio de publicação** - indica o impacto de um trabalho; **riqueza da comunidade** - usada para normalizar os valores aferidos com relação ao poder econômico das comunidades.

2.1.1.1. Problemática

Apesar dos esforços em se estabelecer normas matemáticas e estatísticas para valorar a produção científica, a ciencimetria ainda é olhada com desconfiança e medo. Pesquisadores temem ver suas linhas de pesquisa serem desvalorizadas por números que, na verdade, podem não refletir uma realidade. Instituições temem perder status e investimentos, governos temem que suas políticas de desenvolvimento científico e educacional sejam avaliadas pelo público [8]. Como já foi dito, medir o fruto do intelecto não é uma tarefa fácil.

A falta de um agente centralizador de informações dos trabalhos em C&T também é um grave problema. Geralmente, os textos estão espalhados em diversas bases de dados, que por interesses de ordem econômica não são tornadas acessíveis a todo o público interessado. Ainda quando existe a convergência de aceitação, a análise dos dados é complexa, devido ao seu volume e aos seus relacionamentos. O uso da computação tem auxiliado especialistas na análise dos dados cienciométricos. Tecnologias como mineração de dados e visualização de informações são empregadas na ciencimetria.

2.1.2. Visualização de Informações e Aplicações em Ciencimetria

Visualização de Informações é uma área emergente de pesquisa que busca formas inovadoras de se ampliar a legibilidade dos mais diversos tipos de dados abstratos. Por meio de suas técnicas, dados geralmente sem geometria implícita, e comumente estudados em forma de tabelas, são associados a representações visuais interativas, enfatizando a sua estrutura interna e auxiliando na percepção de padrões.

Knight e outros [7] consideram que uma boa visualização deve apresentar duas características. A primeira é a **expressividade** - adequação da mesma à tarefa a qual ela se propõe auxiliar, apresentando apenas os dados necessários para

entender o que se quer transmitir e nada mais. A segunda característica é a **efetividade** - uma visualização efetiva é aquela que favorece a compreensão do modelo, buscando mapeamentos intuitivos dos dados em relação ao escopo dos especialistas, [3] e que possibilita alta interatividade com os dados, explorando ao máximo o potencial computacional disponível [1].

Algumas técnicas de visualizações de informações já vêm sendo empregadas com sucesso para auxiliar nas atividades de Cienciometria. A Figura 2 exemplifica a redução de várias dimensões em duas e mostra como os dados podem ser sobrepostos sem afetar a sua compreensão.

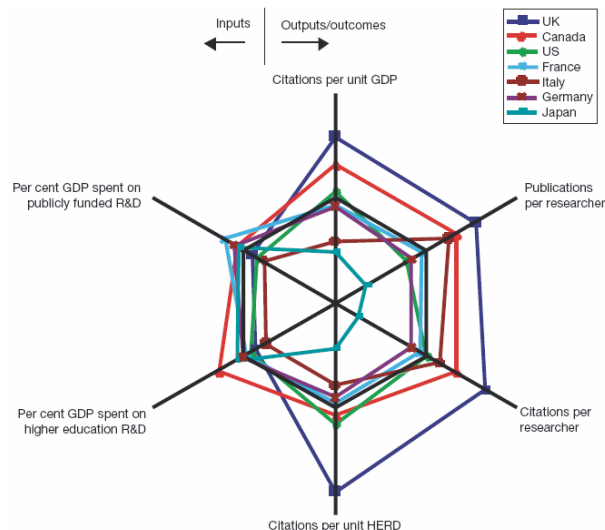


Figura 2. Comparação dos investimentos financeiros (*incomes*) com os resultados científicos (*outcome*). Figura extraída do artigo *The scientific impact of nations* [6]. Uma visualização produzida a partir de dados cienciométricos e mostra o grau de pesquisa das nações G8 (grupo dos oito países mais ricos do mundo). Os dados foram normalizados pela média econômica das sete nações, sendo representados em um hexágono simétrico (técnica conhecida como *starplot*). GDP (*Gross Domestic Public*) é o mesmo que PIB (Produto Interno Bruto); R&D (*Research and Development*) é o mesmo que P&D (Pesquisa e Desenvolvimento); HERD (*Higer Education Funding of R&D*) indica o capital de investimento da educação superior em P&D.

2.1.2.2. Mineração Visual de Dados: Visão além do alcance

Assim como a mineração de dados busca evidenciar fatos e relações ocultas nas bases de dados convencionais [4][5], uma boa visualização também pode revelar fatos e relações ocultas nos resultados gerados até mesmo pela própria mineração de dados.

Nesse contexto, a mineração visual de dados busca oferecer soluções gráficas enfatizando relações e padrões nos dados não necessariamente advindos de processos de mineração de dados, os quais são computacionalmente complexos.

2.1.3 A Plataforma Lattes

A Plataforma Lattes é de grande relevância para a produção técnico-científica brasileira, e ganhou nos últimos anos uma projeção internacional, através de parcerias estabelecidas com países da América e Europa, formando a *Rede ScienTI* [9].

Buscando promover a socialização dos dados contidos na Plataforma Lattes, tornando possível a integração de suas bases de dados com os mais

variados sistemas utilizados pelas instituições de pesquisa, foi desenvolvida uma ontologia chamada **LMPL** ou **Linguagem de Marcação da Plataforma Lattes**. Esta linguagem utiliza o padrão XML para definir significado aos dados da plataforma. Assim, os sistemas podem interoperar utilizando estes padrões, sem necessariamente compartilharem a mesma estrutura de armazenamento de dados. [9].

2.2. Abordagem

Como já foi mencionado, um dos problemas da representação visual de dados é quanto a sua multidimensionalidade. Atualmente a Plataforma Lattes dispõe de vinte e oito indicadores de produção científica, ou seja, vinte e oito dimensões.

Os dados foram extraídos de uma coleção de currículos no formato LMPL, cedidos através do convênio Extrator Lattes. Estes dados foram convertidos em um banco de dados relacional, para facilitar as rotinas de consultas. A partir daí, cada indicador de produção foi mapeado para um eixo paralelo devidamente escalonado pelo maior valor de produção do universo de dados. Os dados de produção de cada pesquisador foram mapeados para as linhas horizontais, que indicam o valor atingido por cada pesquisador em cada eixo.

4. RESULTADOS E DISCUSSÃO

Após o mapeamento, foi criada uma ferramenta de visualização chamada VLattes (Visual Lattes), que implementa uma técnica de coordenadas paralelas. Um problema já conhecido desta técnica é a sobreposição das linhas. Para atenuar esta deficiência, o VLattes imprime em cada marca de valor uma “pegada”, que se dirige para o vermelho a cada vez que uma linha passa por um determinado valor. Desta forma, a cada vez que uma linha passa por alguma marca ela deixa o seu “rastro”.

O usuário pode interagir com o VLattes de três formas: a primeira é criando restrições de valores em cada barra. Desta forma as linhas que não corresponderem ao conjunto de restrições serão desativadas (colocadas em cor diferente ou simplesmente ocultadas, dependendo das predefinições do usuário). A segunda forma é definir pesos para cada indicador de produção, causando um redimensionamento das barras com base no novo peso especificado. Para isso basta que o usuário defina um novo peso através do campo localizado abaixo de cada eixo. Esta ação não só altera a proporcionalidade visual das barras, como também altera o score geral dos pesquisadores, exibido ao lado do nome de cada pesquisador, no eixo de pesquisadores. Este score é calculado através da média ponderada de cada indicador de produção e seu respectivo peso. A terceira forma de se interagir com o VLattes é ativando a lente de aumento, disposta na parte mais a direita dos eixos. Inicialmente, esta lente somente aumenta os nomes da lista de pesquisadores à medida que é deslizada. Mas o usuário pode ativá-la para mostrar somente os dados de produção dos pesquisadores cujos nomes estão sendo atualmente ampliados. Este recurso é útil para amenizar o outro grande problema da técnica de coordenadas paralelas: a densidade de linhas exibidas quando ao número de pesquisadores crescem.

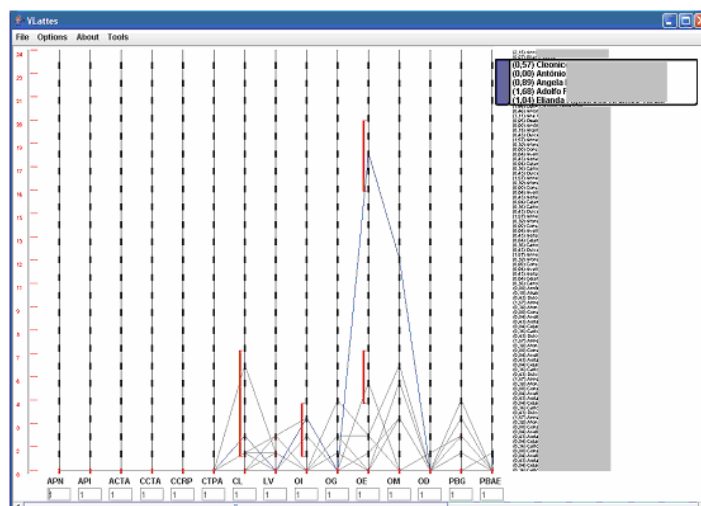


Figura 3. Screenshots da ferramenta VLattes

4. CONCLUSÃO

A linguagem LMPL abriu novos horizontes para a pesquisa exploratória na área da cienciometria nacional. A definição de uma visualização capaz de responder com mais eficiência as questões levantadas por técnicos e usuários da Plataforma Lattes representa uma importante ferramenta para o suporte a tomada de decisão. As instituições e pesquisadores poderão fazer uma análise mais crítica de suas atuações no contexto nacional e, paralelamente, definir objetivos e metas para corrigir as deficiências existentes.

Muitos problemas ainda devem ser solucionados, e a cada dia novas questões cienciométricas surgem, fato que amplia o desafio da construção de um conjunto de visualizações que possibilitem a mineração visual de dados.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Ball, T. and Eick, S. G. (1996). Software visualization in the large. IEEE Computer, pages 33-43.
- [2] Boyack, K. W., Wylie, B. N., and Davidson, G. S. (2001). Domain visualization using vxinsight R for science and technology management. Journal of the American Society for Information Science and Technology (JASIST).
- [3] Carla M. Dal Sasso Freitas, e. a. Evaluating usability of information visualization techniques.
- [4] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). Knowledge discovery and data mining: Towards a unifying framework. AAAI Press.
- [5] Han, J. and Kamber, M. (2001). Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco, CA.
- [6] King, D. A. (2004). The scientific impact of nations. Nature Publishing Group, 430:331-316.
- [7] Knight, C. (2001). Visualisation effectiveness. Conference on Imaging Science, Systems, and Technology.
- [8] Macias-Chapula, C. A. (1998). O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. Ci. Inf., 27(2):134-140.
- [9] Pacheco, R. C. S. and Kern, V. M. (2001). Uma ontologia comum para a integração de bases de informações e conhecimento sobre ciência e tecnologia. Ciência da Informação, Brasília, 30(3):56-63.