



# Perfil de Produção Bibliográfica dos Programas Brasileiros de Pós-Graduação em Ciência da Computação

Jesús P. Mena-Chalco<sup>1</sup>, Luciano A. Digiampietri<sup>2</sup>, Leonardo B. Oliveira<sup>3</sup>

<sup>1</sup>Centro de Matemática, Computação e Cognição da Universidade Federal do ABC

<sup>2</sup>Escola de Artes, Ciências e Humanidades da Universidade de São Paulo

<sup>3</sup>Instituto de Ciências Exatas da Universidade Federal de Minas Gerais

Eixo temático: Produção e Produtividade Científica

Modalidade: Apresentação oral

## RESUMO

A área de Ciência da Computação tem se caracterizado por ser relativamente nova e dinâmica, atuando de forma interdisciplinar com diferentes áreas do conhecimento. No Brasil, atualmente existem mais de 45 programas acadêmicos de pós-graduação que vem tendo um rápido crescimento em termos de produção intelectual. Diferentes são as formas de caracterizar as áreas de pesquisa, sendo uma delas o perfil ou forma de produção bibliográfica. Nesse sentido, este trabalho apresenta uma descrição do perfil de produção bibliográfica dos Programas de Ciência da Computação avaliados pela CAPES nos triênios 2004-2006 e 2007-2009. Como resultado da análise do perfil podemos destacar que a área de Computação Brasileira se caracteriza por publicar preferencialmente trabalhos completos em anais de congressos (aproximadamente 54%), seguido de artigos completos em periódicos (aproximadamente 14%).

**Palavras-chave:** perfil de produção bibliográfica; artigos publicados em anais; produção em ciência da computação.

## 1 INTRODUÇÃO

A área de Ciência da Computação criou uma cultura científica diferente, com processos, métodos e hábitos diferentes no atuar científico devido ao amplo uso de tecnologia, no qual mudanças ocorrem de maneira muito rápida. A Ciência da Computação tem um modelo de atuação diferente das outras áreas.

Neste modelo os trabalhos publicados em conferências científicas são bastante valorizados e o processo de revisão é bastante parecido com as revisões de artigos em revista, contando, por exemplo, com múltiplos revisores para cada artigo e revisão as cegas. Porém o

peso real que é dado aos artigos publicados em anais é bastante controverso (FRANCESCHET, 2010; FREYNE et al., 2010).

Entender este perfil é fundamental para que revisões na literatura não ignorem artigos relevantes, cujos conteúdos estejam publicados apenas nos anais de conferências científicas. Além disso, publicações de livros, ao contrário do que ocorrem em diversas outras áreas, não têm tanto peso científico na Ciência da Computação (FRANCESCHET, 2010; FREYNE et al., 2010).

No Brasil, o comitê da área de Ciência da Computação da CAPES atribui um classificador *Qualis* aos eventos científicos e não apenas aos periódicos científicos, por considerar a relevância desse tipo de evento para a divulgação de pesquisas de ponta na área. A atribuição de um dado estrato do *Qualis* para as conferências é baseada no número médio de citações de cada artigo publicado nos anais de um dado evento (metodologia semelhante ao cálculo do fator de impacto de uma revista).

O número de programas de pós-graduação na Área de Ciência da Computação no Brasil mais que triplicou nos últimos 15 anos (MODESTO, 2010) e a produção científica na área cresceu amplamente. O objetivo deste trabalho é descrever o perfil da produção bibliográfica dos programas brasileiros de pós-graduação em Ciência da Computação. Para a análise dos dados, foram considerados todos os programas acadêmicos de pós-graduação avaliados pela CAPES no triênio 2004-2006 e 2007-2009.

## **2 ESTRATÉGIA DE IDENTIFICAÇÃO DO PERFIL DE PRODUÇÃO**

Neste trabalho consideramos como única fonte de dados as publicações cadastradas na Plataforma Lattes, pois atualmente a Plataforma é um sistema (padrão) nacional onde são registradas informações acadêmicas e profissionais dos pesquisadores atuantes na ciência brasileira (AMORIN, 2003).

Para a identificação do perfil de produção bibliográfica da área de Ciência da Computação foi adotada a seguinte estratégia composta de 4 fases: (i) identificação dos programas de pós-graduação avaliados pela CAPES, (ii) identificação dos professores, na modalidade de *professor permanente*, associados a cada programa de pós-graduação, (iii) identificação dos currículos Lattes de cada professor associado a cada programa, e finalmente (iv) extração das publicações bibliográficas de cada programa.

## 2.1 FASE 1

Os programas de pós-graduação avaliados nos triênios 2004-2006 e 2007-2009 foram extraídos dos *Documentos da Área da CAPES*<sup>1</sup> correspondentes à Área de Ciência da Computação. Ao todo, foram identificados 39 programas de pós-graduação acadêmicos avaliados nos triênio 2004-2006 e 45 programas acadêmicos avaliados no triênio 2007-2009.

## 2.2 FASE 2

Para cada triênio e cada programa de pós-graduação foram identificados os nomes completos de todos os professores, na modalidade de *permanentes*, registrados nos *Cadernos de Indicadores da CAPES*<sup>2</sup>. Ao todo, foram identificados 753 professores correspondentes ao triênio 2004-2006, e 889 professores correspondentes ao triênio 2007-2009.

## 2.3 FASE 3

Na Plataforma Lattes, todo currículo é acessível via web através de um código único de acesso composto de 16 algarismos (denominado ID Lattes). Nesta fase foram obtidos todos os códigos Lattes de todos os professores identificados na fase anterior. O processo de identificação de currículos Lattes foi realizado de forma semiautomática, através de consultas web dos nomes completos dos professores nos buscadores Google e Bing. É importante ressaltar que foi realizada uma validação/verificação manual sobre todos os IDs Lattes identificados a fim de garantir a correta associação dos currículos na coleta dos dados.

## 2.4 FASE 4

Para a extração das produções bibliográficas de cada programa de pós-graduação foi utilizado o scriptLattes (MENA-CHALCO; CESAR-JR, 2009). Esta ferramenta de software-livre, desenvolvido no IME/USP permite criar relatórios de produção acadêmica de diferentes grupos de pesquisadores cadastrados na Plataforma Lattes. Na sumarização dos dados, todas as produções científicas são discretizadas por tipos (por exemplo, Artigos completos publicados em periódicos, ou Capítulos de livros). Para cada ano, todas as publicações pertencentes a produções realizadas em coautoria são automaticamente identificadas, através das comparações entre os títulos das publicações. Duas publicações são consideradas iguais se

---

<sup>1</sup> Os Documentos da Área da CAPES para a área de Ciência da Computação estão disponíveis em [http://capes.gov.br/images/stories/download/avaliacao/COMPUTACAO\\_05mar10.pdf](http://capes.gov.br/images/stories/download/avaliacao/COMPUTACAO_05mar10.pdf)

<sup>2</sup> Os Cadernos de Indicadores da CAPES para todas as áreas estão disponíveis em <http://conteudoweb.capes.gov.br/conteudoweb/CadernoAvaliacaoServlet>

seus títulos forem pelo menos 90% semelhantes (NAVARRO, 2001). Essa abordagem permite obter uma boa aproximação das quantidades reais de produções publicadas durante um determinado período (e.g. triênio).

Nesta fase, tendo definida a lista de currículos Lattes de todos os professores associados aos programas de pós-graduação, foi projetado um *script* que permite executar de forma automática o scriptLattes para diversos grupos de pesquisa. Ao todo, o scriptLattes foi executada 86 vezes, correspondentes a 40 grupos (39 programas + o conjunto total de professores) referentes ao triênio 2004-2006 e 46 grupos referentes ao triênio 2007-2009 (45 programas + o conjunto total de professores).

### 3 PERFIL DE PRODUÇÃO BIBLIOGRÁFICA

Neste trabalho, as análises foram realizadas em duas partes correspondentes à identificação do perfil de produção bibliográfica (i) de cada programa de pós-graduação, e (ii) do conjunto total de professores associados aos programas de pós-graduação em Ciência da Computação.

#### 3.1 PERFIL POR PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

Na Figura 1, apresenta-se a produção bibliográfica obtida para cada programa de pós-graduação em Ciência da Computação. Observe que, nas tabelas, cada linha está associada a um tipo de produção bibliográfica, e cada coluna ( $P_i$ ) a um programa de pós-graduação. As barras coloridas representam as porcentagens de produções bibliográficas de cada programa de pós-graduação (cada cor está associada um tipo de produção bibliográfica). Da figura podemos destacar que, para os dois triênios, a grande maioria os programas de pós-graduação publicam artigos completos em congressos/eventos.

É possível também observar que o perfil das publicações quase não mudou de um triênio para o outro. Enquanto várias áreas estão se esforçando para concentrar suas publicações em periódicos, na Ciência da Computação este esforço é sentido ainda de maneira muito sutil e a maioria dos artigos continua sendo publicada em anais de eventos e as pequenas diminuições relativas no número de artigo completos publicados em anais de congressos no triênio 2007-2009 se dá pelo aumento de resumos expandidos publicados em anais de congressos no mesmo período.

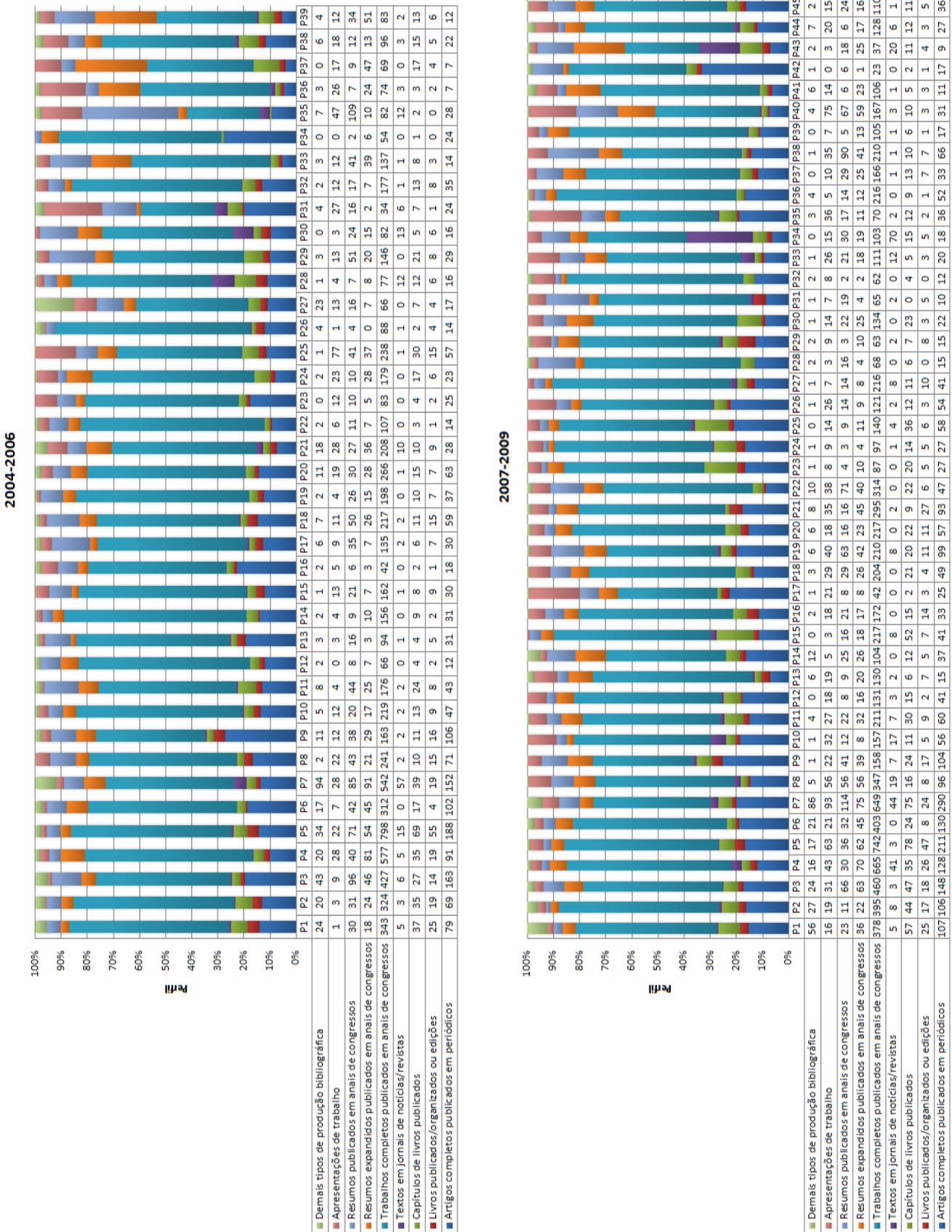
### 3.2 PERFIL DOS PROFESSORES DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

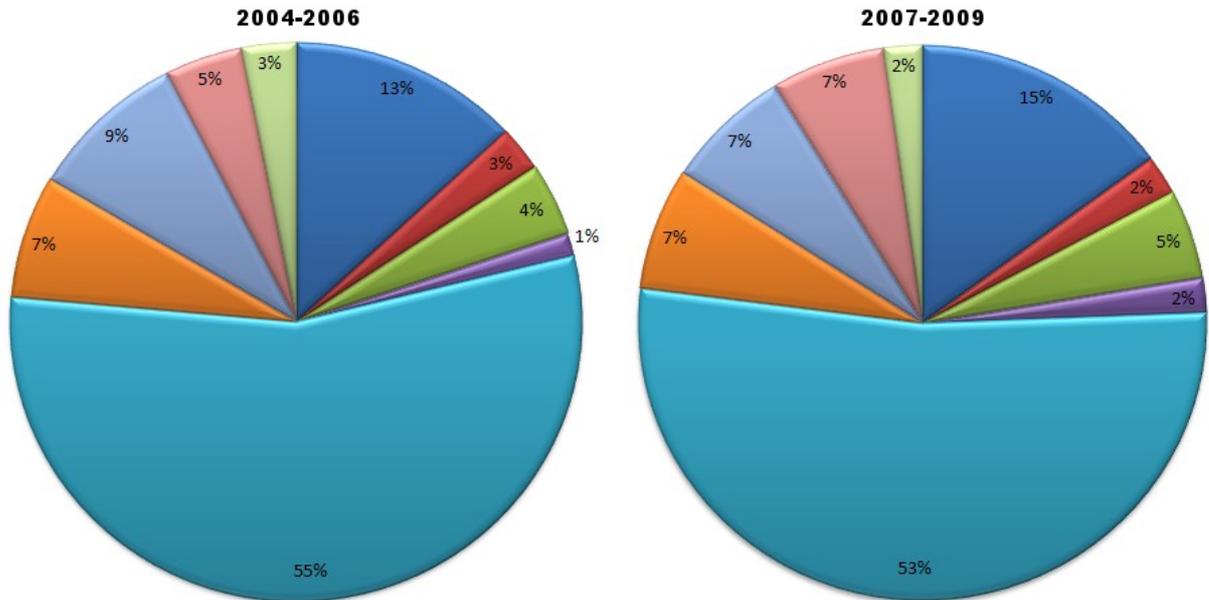
Na Figura 2, apresenta-se a produção bibliográfica obtida para o conjunto total de professores associados a todos os programas de pós-graduação em Ciência da Computação. É importante lembrar que foram considerados 753 e 889 professores dos programas avaliados nos triênios 2004-2006 e 2007-2009, respectivamente. Observe que para os dois triênios, a área de Ciência da Computação mantém o mesmo perfil, caracterizando-se por publicar preferencialmente trabalhos completos em anais de congressos (aproximadamente 54%), seguido de artigos completos em periódicos (aproximadamente 14%). Note também que os Livros publicados/organizados ou edições representam uma parte pouco expressiva de produção da área (aproximadamente 3%). Finalmente, a área de Ciência da Computação tem a peculiaridade de ter a menor produção bibliográfica, em ambos os triênios, quando se trata de Textos em jornais de notícias/revistas. No segundo triênio analisado houve um pequeno aumento na porcentagem de artigos publicados em periódicos (de 13% para 15%), para os demais tipos de publicação as variações foram menores.

## 4 CONSIDERAÇÕES FINAIS

Em geral, os programas de pós-graduação associados a área de Ciência da Computação dão maior ênfase na publicação de trabalhos completos em congressos/eventos. Esta característica já foi discutida em estudos semelhantes (FRANCESCHET, 2010; FREYNE et al., 2010). Este trabalho descreveu a produção bibliográfica dos programas brasileiros de pós-graduação em Ciência da Computação nos dois últimos triênios (2004-2006 e 2007-2009) avaliados pela CAPES, apresentando a prevalência das publicações em anais de eventos científicas. Porém, neste trabalho também foi mostrado um aumento na produção de periódicos no último triênio, o que pode indicar que os programas da área estão preocupados em acompanhar os esforços de diversas áreas no aumento da publicação de artigos em periódicos.

Como direcionamentos futuros pretendemos utilizar a estratégia descrita neste trabalho para calcular, avaliar e comparar o perfil de produção bibliográfica de diferentes áreas do conhecimento. Trabalhos nessa linha permitirão evidenciar as diferentes formas de atuação acadêmica de cada área acadêmica no Brasil.





	2004-2006	2007-2009
■ Artigos completos publicados em periódicos	1658	2344
■ Livros publicados/organizados ou edições	329	366
■ Capítulos de livros publicados	533	816
■ Textos em jornais de notícias/revistas	169	310
■ Trabalhos completos publicados em anais de congressos	7015	8292
■ Resumos expandidos publicados em anais de congressos	896	1114
■ Resumos publicados em anais de congressos	1146	1140
■ Apresentações de trabalho	558	1022
■ Demais tipos de produção bibliográfica	394	351

**Figura 2.** Perfil de produção bibliográfica do conjunto total de professores associados a todos os Programas Brasileiros de Pós-graduação em Ciência da Computação avaliados nos triênios 2004-2006 e 2007-2009.

## REFERÊNCIAS

- AMORIN, C. V. Curriculum vitae organization: The Lattes software platform. **Pesquisa Odológica Brasileira**, v. 17, n. 1, p. 18–22, 2003.
- FRANCESCHET, M. The role of conference publications in CS. **Communications of ACM**, 53, 129-132, 2010.
- FREYNE, J.; COYLE, L.; SMYTH, B.; CUNNINGHAM, P. Relative status of journal and conference publications in computer science. **Communications of ACM**, 53, 124-132, 2010.
- MENA-CHALCO, J. P.; CESAR-JR, R. M. scriptLattes: An open-source knowledge extraction system from the lattes platform. **Journal of the Brazilian Computer Society**, v. 15, n. 4, p. 31–39, 2009.
- MODESTO, M. A. B. Análise multidimensional da produção científica em ciência da computação. **Tese de Doutorado, UFMG**, 2009.
- NAVARRO, G. A guided tour to approximate string matching. **ACM Computing Surveys**, v. 33, n. 1, p. 31–88, 2001.