



## TENDÊNCIAS DA CIÊNCIA DA INFORMAÇÃO: BIG DATA E INTELIGÊNCIA COMPETITIVA

**Kátia Kelvis Cassiano<sup>1</sup>, Renata Moreira Limiro<sup>2</sup>, Douglas Farias Cordeiro<sup>3</sup>**

<sup>1</sup>*Faculdade de Informação e Comunicação da Universidade Federal de Goiás, katiakelvis@ufg.br*

<sup>2</sup>*Faculdade de Informação e Comunicação da Universidade Federal de Goiás, renatamlimiro@gmail.com*

<sup>3</sup>*Faculdade de Informação e Comunicação da Universidade Federal de Goiás, cordeiro@ufg.br*

**Resumo:** O crescente volume de dados disponíveis associado à necessidade de soluções voltadas à geração de informação e descoberta do conhecimento, com maior acurácia e assertividade, é um dos grandes desafios do fenômeno conhecido como Big Data. E este cenário é ainda mais complexo no âmbito da Inteligência Competitiva, a qual busca alcançar a geração de informação útil e competitiva nos negócios. Tais campos de pesquisas se convergem diretamente dentro da área da Ciência da Informação, enquanto tendências de pesquisa e geração de soluções inovadoras. Entretanto, compreender e utilizar os conhecimentos destas áreas ainda é um desafio e demanda estudos e análises específicos. Diante disso, este artigo se propõe a apresentar um estudo quantitativo acerca das produções científicas da Inteligência Competitiva e Big Data por pesquisadores da Ciência da Informação, através de análises de dados provenientes de currículos Lattes, gerando insumo, por meio de visualizações gráficas, para discussões no contexto destas áreas.

**Palavras-chave:** Ciência da Informação. Tendências. Inteligência Competitiva. Big Data. Análise Descritiva.

### 1. Introdução

A necessidade de exploração da informação como diferencial estratégico é algo fundamental e de grande importância nos mais diversos contextos, e sua geração e disseminação está diretamente ligada à Ciência da Informação, uma área de caráter interdisciplinar que possui como objeto principal a informação. Neste contexto, as evoluções tecnológicas, assim como as mudanças de modelos de negócio, as demandas competitivas, e o constante aumento no volume de dados, acabam por gerar cenários onde se torna essencial o emprego de modelos e técnicas inovadoras, que possibilitem gerar resultados mais assertivos e que agreguem mais valor aos negócios.

Neste sentido, a inovação passou a ser elemento chave na Ciência da Informação, e a identificação de tendências é algo de grande relevância para o crescimento e exploração dessa área. Diante disso, a exploração de modelos e conceitos da Inteligência Competitiva se tornou algo de grande interesse, uma vez que provê mecanismos e estratégias para a obtenção de informações que sejam um diferencial para os negócios. Todo este arcabouço de estratégias está diretamente ligado às capacidades de manipulação e transformação de grandes volumes de dados em informação, no âmbito de um fenômeno denominado Big Data.

A partir disso, se emerge um cenário onde a Inteligência Competitiva e o Big Data podem ser destacados como tendências na área da Ciência da Informação. Assim, prover mecanismos para compreender o estudo e exploração destes temas no âmbito da Ciência da Informação é de grande relevância. Neste sentido, este artigo se propõe a apresentar uma análise descritiva sobre tais áreas a partir de dados obtidos através dos currículos Lattes de pesquisadores da área da Ciência da Informação, obtidos através da ferramenta Script Lattes, apresentando resultados quantitativos que exibem a presença e impacto destes temas em estudos científicos e acadêmicos.

## **2. Ciência da Informação e Tendências**

A Ciência da Informação, de acordo com Souza e Stumpf (2009), pode ser compreendida como uma área interdisciplinar, a qual possui como foco principal a informação, atuando em questões englobam análise, coleta, manipulação, classificação, disseminação e recuperação da informação, promovendo mecanismos para garantir a geração de informação a partir de dados, e a consequente descoberta do conhecimento. Segundo Russo (2010), o marco inicial de surgimento desta área está ligado à publicação de um trabalho, no início o período pós-guerra, em 1945, abordando questões relacionadas ao valor e volume de informação geradas durante a Segunda Guerra Mundial.

Após este marco inicial, a Ciência da Informação se consolidou enquanto área durante os anos 1960, como resposta a demandas informacionais que na altura extrapolavam as possibilidades e competências apresentadas por outras áreas (ARAÚJO, 2014). Esta nova área passou a possibilitar o estudo e implantação de processos característicos, voltados à geração e disseminação de informação, possuindo uma ligação direta com áreas como a Biblioteconomia,

Arquivologia, Museologia, Matemática e, de certa forma, também com a Ciência da Computação, uma vez que a evolução da Ciência da Informação é fortemente relacionada aos avanços tecnológicos possibilidades alcançadas através de métodos e rotinas desenvolvidos no âmbito da Ciência da Computação.

Neste contexto, um dos pontos que possui grande destaque se refere à transformação dos dados em informação útil à geração de conhecimento e tomada de decisão. As evoluções no campo dos Sistemas de Informação e no âmbito da Internet, principalmente, acabaram por gerar um cenário onde a criação de dados é algo comum e crescente. Dados operacionais e transacionais, assim como dados gerados por mídias digitais, passam se configura como um fenômeno denominado de Big Data (GOMES e BRAGA, 2016). Neste sentido, o desafio maior passou a ser a determinação de modelos e métodos que sejam capazes e trabalhar sobre volumes massivos de dados e garantir a geração de informação estratégica, competitiva e diferencial ao negócio.

Esse cenário de grandes volumes de dados se encontra também intrinsecamente ligado às possibilidades de estudos e análises com propósito de obtenção de vantagens competitivas, o que se configura como elemento motriz da Inteligência Competitiva, que surge também como sendo uma área ligada à Ciência da Informação. A partir disso, a utilização e implantação de soluções que se baseiam na aplicação de técnicas provenientes da Estatística, Mineração de Dados e Inteligência Artificial se tornaram necessárias e fundamentais na descoberta de conhecimento, revelando uma relação de proximidade entre Ciência da Informação e tais áreas, e revelando o emprego e exploração destas como tendências de caráter estratégico, e como garantia da informação enquanto ativo organizacional que agregue valor aos negócios.

### **3. Inteligência Competitiva**

Uma das premissas básicas da Inteligência Competitiva (IC) é a geração de vantagens competitivas como diferencial estratégico. Neste sentido, de acordo com Jakobiak (1998), a IC pode ser descrita como uma área que encontra-se baseada em conjuntos de ações coordenadas,

as quais envolvem pesquisa, tratamento e disseminação de informações para as partes interessadas de um negócio. Ainda nesta linha, segundo Kahaner (1996), a IC pode ser compreendida como um processo organizacional que busca analisar e qualificar informações de forma sistemática e inteligente com o objetivo de gerar vantagem e diferencial competitivo. Em suma, pode ser afirmado que o objetivo principal da inteligência competitiva é a geração de vantagens competitivas sobre o mercado. Diante disso, para a contemplação de tal objetivo, é fundamental que uma sequência lógica de procedimentos sejam levados em consideração, se levando em conta também o contexto da organização e sua cultura organizacional.

De acordo com Coelho (1999), a IC pode ser descrita como uma área voltada à geração contínua do conhecimento, com início na coleta de dados, os quais devem então ser submetidos a tratamento e organização e, posteriormente, transformados em informação, a qual, de forma associada a estudos, análises e contextualização é principal insumo para a geração de conhecimento, auxílio à tomada de decisão e provimento de vantagens competitivas. Nesta mesma linha, Prescott (2002), descreve a IC como um conjunto de metodologias voltadas à coleta e análise de forma ética e legal, de informações provenientes de concorrentes, com o propósito de gerar vantagem competitiva. Diante disso, é fundamental destacar que o cenário corporativo apresenta particularidades e características que o tornam relativamente volátil, no qual é comum que produtos, serviços e tendências surjam e desapareçam de forma contínua, assim como ocorre com a atualização dos modelos de negócios, o que potencializa a necessidade de emprego da IC como solução estratégica para as organizações. Nestes cenários, para que um modelo de IC seja desenvolvido e implantado de maneira eficaz é de grande importância que haja adaptação à produção e disseminação de informação, garantindo uma cultura de IC no âmbito organizacional. Essa geração de informação está diretamente ligada à coleta de dados, os quais comumente se tratam de grandes volumes de dados, associados ao que é conhecido como Big Data.

### **3. Big Data**

De acordo com Gomes e Braga (2017), o termo Big Data está relacionado à densos volumes de dados de alta variedade. Neste sentido, com o propósito de geração de informação, é necessário o emprego de técnicas e modelos que tenham como características a inovação e a rentabilidade. Ainda de acordo com as autoras, esse fenômeno é impulsionado principalmente pela evolução tecnológica e consequente diminuição dos custos associados ao armazenamento e processamento de dados. Neste contexto, de acordo com Vance (2011), diversos tipos de aplicações Big Data estão relacionadas a dados transacionais e operacionais de uma

organização, com o propósito de gerarem insumo para a realização de rotinas de melhoria em processos de negócio, desempenho de distribuição de informação e estabelecimento de canais com o cliente.

#### **4. Mineração de Dados**

A busca por padrões válidos, úteis compreensíveis a partir do processamento de grandes bases de dados é um desafio constante em diversas organizações e cenários. Neste contexto, é comum que a capacidade de acumulação de dados seja consideravelmente maior que a de extração de informações, fato este que justifica a demanda por soluções que proporcionem auxílio na descoberta de informações consideradas relevantes, o qual pode utilizar a metodologia conhecida como Descoberta de Conhecimento em Bases de Dados (do inglês, *Knowledge Discovery in Databases - KDD*).

O KDD se refere a um conjunto de etapas que são realizadas com o propósito de buscar e extrair padrões de conhecimento em bases de dados. De acordo com Gouveia (2009), o resultado deste processo se refere a informações anteriormente implícitas e desconhecidas, as quais servirão de auxílio à tomada de decisão. De acordo com Fayyad et al. (1996), o KDD pode ser descrito como um processo não trivial voltado à identificação de novos padrões úteis e compreensíveis. Esse processo engloba a realização cíclica de cinco atividades: seleção, pré-processamento, transformação, mineração e interpretação. A etapa de mineração pode ser destaca como o ponto chave do processo, fundamental para a geração de informação e descoberta do conhecimento, na qual é realizada a determinação das tarefas a serem realizadas (Descrição, Classificação, Regressão, Predição, Agrupamento ou Associação), assim como a execução das rotinas referentes aos algoritmos pertencentes a cada tarefa (FELIPE JÚNIOR, 2012).

Segundo Bueno e Viana (2012), a mineração de dados pode ser descrita como uma técnica voltada à exploração de grandes conjuntos de dados, possuindo como objetivo o estabelecimento de relações, associações e padrões que originalmente possuem uma visualização complexa, proporcionando a transformação de dados brutos em informação de valor. Nesta mesma linha, para Amaral (2016), a mineração de dados é descrita como um conjunto de processos para exploração e análise de grandes volumes de dados, afim de extrair padrões, previsões, erros e associações.

Neste contexto, a Análise Descritiva se caracteriza como o estudo dos dados coletados com o propósito de geração de informação e apresentação de possíveis cenários anteriormente não visualizados, de forma a realizar uma descrição dos aspectos e características mais

importantes na amostra considerada, conforme descrito por Reis e Reis (2002). De acordo com Cortês, Porcaro e Lifschitz (2002), a Análise Descritiva pode ser descrita como uma subárea da Mineração de Dados responsável pela investigação da representação e descrição de fatos relevantes, desconhecidos e não-triviais em conjuntos de dados, podendo ser ainda dividida em duas etapas: análise prévia e descobrimento.

A análise prévia, no âmbito da Análise Descritiva, se refere à identificação de problemas ou anomalias que apresentam potencial de interferência nos resultados posteriores à Mineração de Dados. Por outro lado, a etapa de descobrimento trata do processo de análise de dado, onde são aplicadas técnicas e rotinas para a busca de padrões ocultos. No contexto do presente trabalho, serão aplicadas técnicas de Análise Descrita voltadas para a efetivação das análises então propostas.

## 5. Web Scraping

A quantidade de informações disponíveis e acessíveis através da Web é significativamente considerável, sendo, inclusive considerada como a maior responsável pelo o que é conhecido como Big Data. Entretanto o acesso e exploração de tais informações de maneira automatizada e massiva pode ser um grande obstáculo na construção de determinados tipos de análises. Diante disso, surge a necessidade de utilização de soluções que proporcionem mecanismos para a extração de dados da Web.

Uma das possíveis soluções é a utilização de APIs (do inglês, *Application Programming Interface*), que de um modo geral podem ser descritas como um conjunto de padrões e rotinas de programação que permitem o acesso a um conjunto de dados ou uma aplicação através da Web (Stair e Reynolds, 2016). Entretanto, é importante destacar que nem sempre os provedores de informação possuem a disponibilidade de serviços de API para extração de seus dados.

Neste cenário, torna-se necessário a exploração de meios alternativos que possibilitem a coleta de dados de forma automatizada. O Web Scraping, conhecido também como Raspagem de Dados, se refere a um conjunto de técnicas computacionais que permitem a extração de dados em páginas Web através da realização de análises de códigos HTML e outros arquivos (MITCHELL, 2015), aplicando rotinas de estruturação destes códigos. Através do Web Scraping é possível recuperar e estruturar dados provenientes da Internet, possibilitando a realização de projetos de análise de dados sob as mais distintas perspectivas.

## 6. Metodologia

O presente trabalho trata de realização de análises descritivas, de caráter quantitativo, sobre dados provenientes de currículos Lattes de pesquisadores da área de Ciência da Informação, disponibilizados pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico do Ministério da Ciência, Tecnologia e Inovação no Brasil, com o propósito de levantar informações sobre a realização de trabalhos relacionados à Inteligência Competitiva e Big Data.

Como parte da fase de pré-processamento, foram identificados os pesquisadores da área de conhecimento Ciência da Informação. Como resultado, uma lista de identificadores foi gerada, os quais constituem a chave de busca de currículos na Plataforma Lattes<sup>1</sup>. Cada currículo Lattes foi obtido por meio da chave de identificação, sendo os dados de cada um dos pesquisadores armazenados em arquivo em formato html (do inglês, *Hypertext Markup Language*).

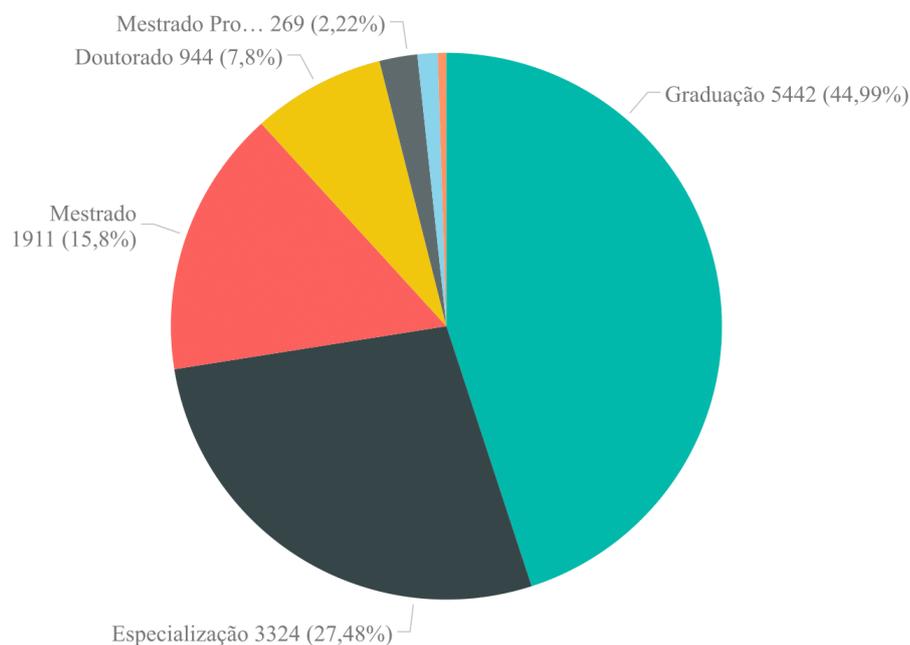
A partir disso, para a realização da análise exploratória, foi desenvolvido um *script* em linguagem de programação Python, utilizando técnicas de mineração de textos. Um corpus textual foi gerado por meio da integração dos arquivos de dados e, de uma forma geral, o script implementado extrai padrões e agrupa os pesquisadores por áreas de atuação, tendo em vista os temas abordados em projetos e trabalhos publicados. Como resultados, tem-se a distribuição dos pesquisadores no Brasil segundo as áreas de interesse, sendo então possível identificar tendências na atuação desses profissionais em relação aos temas de interesse Big Data e Inteligência Competitiva. Os dados foram então categorizados e submetidos a análises, possibilitando a geração de visualizações através de gráficos, os quais apresentam um panorama das pesquisas em tais áreas.

## 7. Resultados e Conclusão

A partir do processamento dos dados obtidos através dos currículos Lattes dos pesquisadores da área de Ciência da Informação, foram geradas uma série de análises descritivas, relacionadas a fatores quantitativos como o nível de titulação, a produção na área de Inteligência Competitiva e Big Data, entre outros. A Figura 1 apresenta a distribuição dos pesquisadores que atuam no Brasil, da área Ciência da Informação, em relação ao nível de instrução.

---

<sup>1</sup> A Plataforma Lattes é uma plataforma virtual criada e mantida pelo CNPq, pela qual integra as bases de dados de currículos, grupos de pesquisa e instituições, das áreas de Ciência e Tecnologia e que atuam no Brasil, em um único sistema de informações.



**Figura 1. Percentual de pesquisadores por nível de formação.**

Observa-se que aproximadamente 45% dos pesquisadores da área Ciência da Informação no Brasil tem formação em nível de graduação. Para o escopo do referido trabalho, optou-se por conduzir a análise com foco nos pesquisadores doutores, sob a justificativa de que os trabalhos registrados nos currículos dos mesmos contemplariam os trabalhos dos pesquisadores de níveis de formação inferiores (especialização, mestrado e graduação).

A Tabela 1 apresenta um resumo das informações relacionadas à produção científica dos pesquisadores doutores no período de 2003 a 2018, com base no processamento dos currículos Lattes dos mesmos na ferramenta ScriptLattes<sup>2</sup>.

**Tabela 1: Produção Científica dos Pesquisadores Doutores.**

Quantidade de pesquisadores doutores	944
Artigos completos publicados em periódicos	11272
Trabalhos completos publicados em anais de congressos	12395
Resumos expandidos publicados em anais de congressos	1738
Resumos publicados em anais de congressos	4348
Total de produção bibliográfica	59704

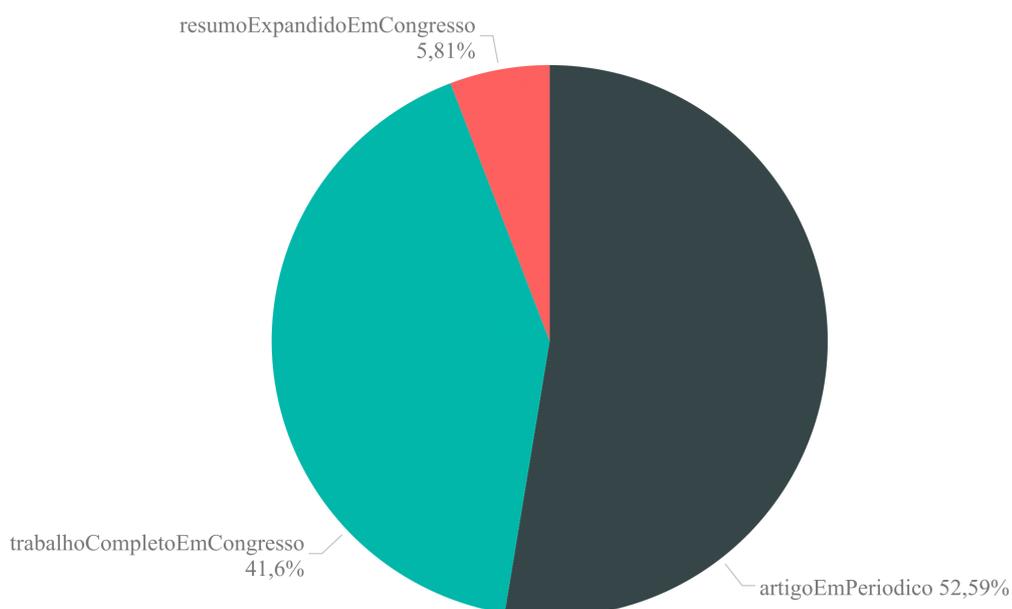
A partir destes resultados, foi gerado um arquivo csv (do inglês, *Comma Separated Values*) contendo dados do processamento composto por 28602 trabalhos, publicados no período de 2003 a 2018. A partir deste arquivo, foram extraídos padrões característicos por

<sup>2</sup> Script Lattes: <http://scriptlattes.sourceforge.net/>

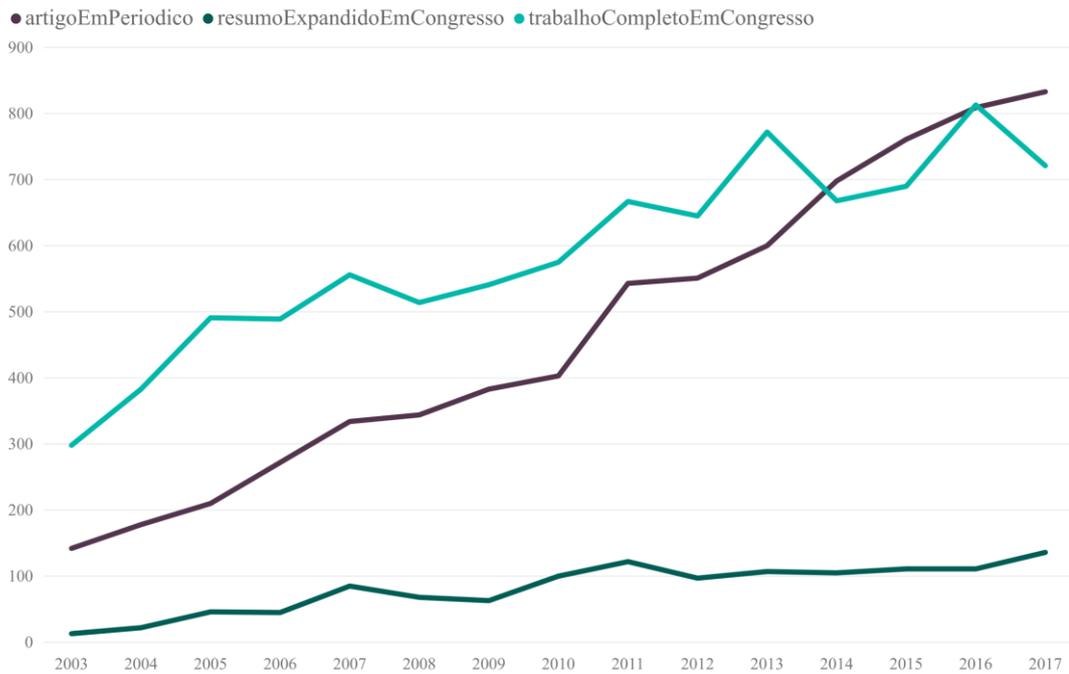
meio de análise exploratória de dados. As Figura 2 e Figura 3 apresentam, respectivamente, a produção dos pesquisadores por tipo de publicação (resumo expandido e trabalho completo em congressos científicos ou artigos científicos publicados em periódicos) e a evolução no período considerado.

Os resultados apresentados demonstram crescimento na produção em Ciência da Informação no período considerado prevalecendo, na média geral, as publicações de artigos científicos em periódicos. As Figura 4 e Figura 5 apresentam uma análise quanto à utilização dos termos Inteligência Competitiva e Big Data, respectivamente.

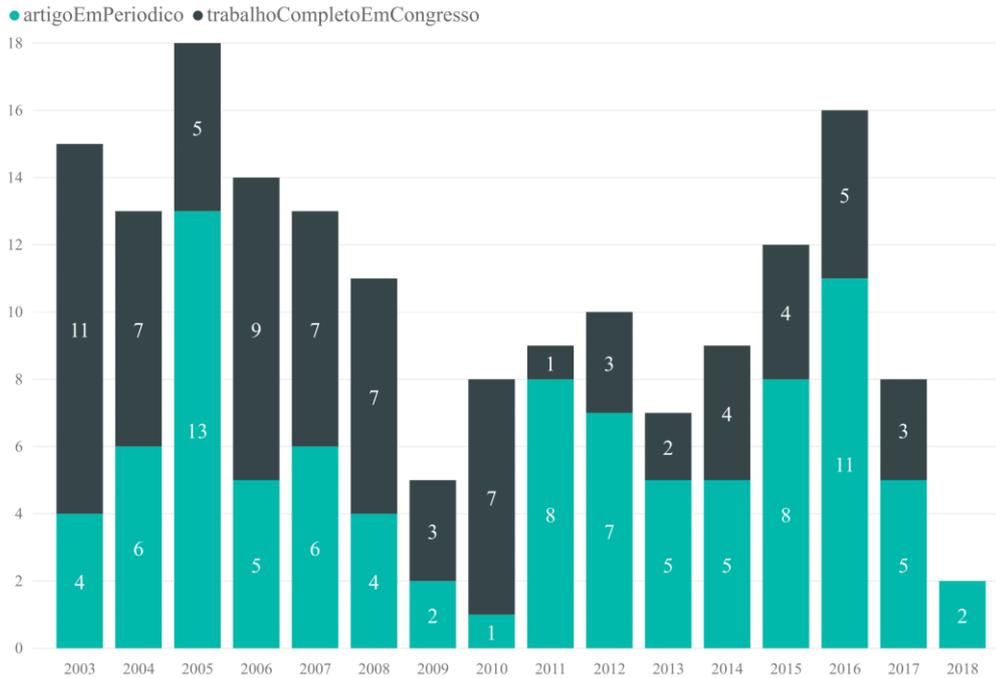
Os resultados apresentados possibilitaram o conhecimento da abrangência da Ciência da Informação em termos do contexto atual da gestão da informação: Big Data e Inteligência Competitiva. É importante observar que, no período de análise, o termo Big Data é abordado nas publicações apenas a partir de 2013, confirmando a motivação inicial do estudo em pauta – tais termos representam uma tendência da atualidade.



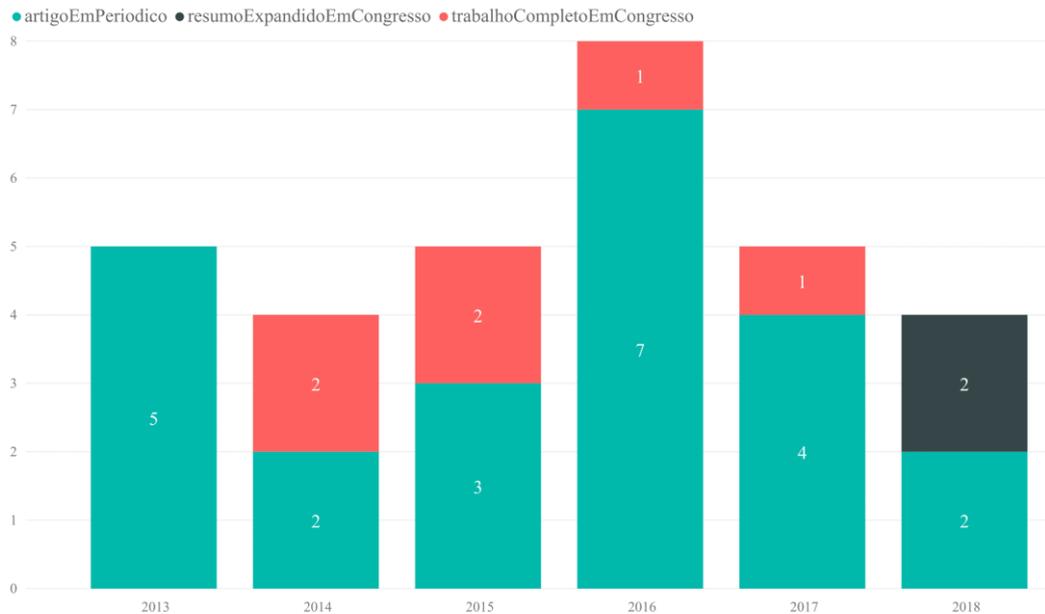
**Figura 2. Percentual de pesquisadores por nível de formação.**



**Figura 3. Produção por tipo no período de 2003 a 2018.**



**Figura 4. Produções científicas que abordam o tema Inteligência Competitiva.**



**Figura 5. Produções científicas que abordam o tema Big Data.**

No Brasil, verifica-se que esta tendência tem respaldo no fato de que tais termos serem tratados, inclusive, no âmbito das estruturas curriculares dos cursos de formação, conforme relatado em outros trabalhos (CORDEIRO e CASSIANO, 2018). Além, disso, é pertinente ressaltar que Big Data e Inteligência Competitiva corroboram com o papel intrínseco da Ciência da Informação no que tange à produção de conhecimento a partir de uma massa de dados, provendo subsídios para a tomada de decisão e garantia da vantagem competitiva no mercado.

## Referências

- Amaral, F. (2016) *Aprenda Mineração de Dados: teoria e prática*. Rio de Janeiro: Alta Books.
- Araújo, C. A. A. (2014) O que é Ciência da Informação? *Informação & Informação*, v. 19, n. 1, p. 1-30.
- Bueno, M. F., & Viana, M. R. Mineração de dados: aplicações, eficiência e usabilidade. In: Congresso de Iniciação Científica do INATEL. Anais... Santa Rita do Sapucaí: INATEL, p. 86-94.
- Cordeiro, F.D., Cassiano, K. K. (2018) Um panorama do ensino superior em gestão da informação no Brasil. *Perspectivas em Ciência da Informação*. No prelo.
- Coelho, G. M. (1999) *Inteligência Competitivo: definindo a ferramenta e seu uso no Brasil*. São Paulo: IBC.
- Cortês, S. C., Porcaro, R. M., & Lifschitz, S. (2002) *Mineração de Dados – Funcionalidades, Técnicas e Abordagens*. Relatório Técnico – Pontifícia Universidade Católica do Rio de Janeiro.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996) From data mining to knowledge discovery in databases. *AI Magazine*, v. 17, n. 3, p. 37-54.

- Felipe Júnior, J. (2012) *Mineração de dados para detecção de fraudes em transações eletrônicas*. Dissertação de Mestrado – Universidade Federal de Minas Gerais, Belo Horizonte.
- Gomes, E., & Braga, F. (2016) *Inteligência Competitiva em Tempos de Big Data - Analisando informações e identificando tendências em tempo real*. Rio de Janeiro: Alta Books.
- Gouveia, R. M. M. (2009) *Mineração de dados em Data Warehouse para sistemas de abastecimento de água*. Dissertação de Mestrado – Universidade Federal da Paraíba.
- Jakobiak, F. (1998) *L'intelligence économique en pratique*. Editions d'Organisations.
- Kahaner, L. (1996) *Competitive intelligence: how to gather, analyze, and use information to move your business to the top*. New York: Touchstone.
- Mitchell, R. (2015) *Web Scraping com Python*. São Paulo: Novatec.
- Prescott, J. E., & Miller, S. H. (2002) *Inteligência competitiva na prática: técnicas e práticas bem sucedidas para conquistar mercados*. Rio de Janeiro: Campus.
- Reis, E., & Reis, I. A. (2002) *Análise Descritiva de Dados*.
- Russo, M. (2010) *Fundamentos de biblioteconomia e ciência da informação*. Rio de Janeiro: E-papes.
- Souza, R. F., & Stumpf, I. R. C. (2009) Ciência da Informação como área do conhecimento: abordagem no contexto da pesquisa e da Pós-Graduação no Brasil. *Perspect. Ciênc. Inf.*, vol.14, n.spe, p.41-58.
- Vance, A. (2011) The data knows. *Bloomberg Businessweek*, p. 70-74.
- Stair, R. M., & Reynolds, G. W. (2016) *Princípios de Sistemas de Informação*. 11 ed. São Paulo: Cengage Learning.