

Mineração de texto aplicada às publicações científicas sobre gestão do conhecimento no período de 2003 a 2012

Cícero Aparecido Bezerra

**Professor do Programa de Pós Graduação em
Ciência, Gestão e Tecnologia da Informação da
Universidade Federal do Paraná**

André José Ribeiro Guimarães

**Mestrando do Programa de Pós Graduação em
Ciência, Gestão e Tecnologia da Informação da
Universidade Federal do Paraná**

<http://dx.doi.org/10.1590/1981-5344/1834>

Desde sua apresentação, em meados da década de 1980, o tema 'Gestão do Conhecimento' (GC) tem sido abordado sob os mais diversos aspectos. Neste sentido, uma vez observado o aumento da quantidade de trabalhos publicados em bases acadêmicas, nos últimos 10 anos, os objetivos da presente pesquisa são (1) encontrar os termos mais empregados nestes trabalhos e as (2) diferenças entre os termos encontrados em publicações brasileiras e estrangeiras. Para isto, empregou-se a técnica de mineração de textos. Os resultados mostram que temas envolvendo práticas, processos, sistemas de GC e tecnologias de informação são recorrentemente encontrados tanto em resumos de artigos nacionais como internacionais. Por outro lado, de maneira geral, os artigos nacionais se distinguem pelo fato de apresentarem termos mais associados a aspectos abstratos da GC (modelos, conhecimento tácito e organizacional), enquanto que os internacionais caracterizam-se por termos mais pragmáticos (efetividade, implementação, relacionamentos e desenvolvimento).

Palavras-chave: *Gestão do conhecimento; Mineração de textos.*

Text mining applied to scientific publications on knowledge management in the period 2003-2012

Since its introduction in the mid-1980s, the theme 'Knowledge Management' (KM) has been approached under different aspects. In this sense, since the observed increase in the quantity of papers published in academic databases in the last 10 years, the objectives of this research are (1) find the terms more used in these works, and (2) differences between the terms found in Brazilian and foreign publications. For this, was applied the text mining technique. The results show that themes involving practices, processes, KM systems and information technologies are recurrently found in both article summaries national and international. Furthermore, in general, national articles are distinguished by the fact that they are associated with a more abstract aspects of the KM (models, tacit and organizational knowledge), while international are characterized by a more pragmatic terms (effectiveness, implementation, relationships and development).

Keywords: Knowledge management; Text mining.

Recebido em 20.06.2013 Aceito em 14.10.2013

1 Introdução

Desde o início da década de 1990, a Gestão do Conhecimento (GC) se tornou um tema de interesse para pesquisadores e consultores de negócios. Este interesse tem se traduzido em artigos, que têm abordado a GC nos seus mais diversos aspectos, dentre eles, capital social (LEE; SUKOCO, 2007, p. 549), inovação (LUNDVALL; NIELSEN, 2007, p. 207), gestão de riscos (PERROTT, 2007, p. 523), impactos organizacionais (SHANG; LIN, 2009, p. 509) e incerteza (De NOMI; ORSI; PILOTTI, 2009, p. 96). No Brasil, os trabalhos publicados na área também têm abordado temas variados, como comunicação científica (LEITE; COSTA, 2007, p. 92), capital intelectual (CAMPOS, 2007, p. 104), inteligência competitiva (CAPUANO; CASAES; COSTA, 2009, p. 19), estrutura organizacional (GONZALEZ; MARTINS; TOLEDO, 2009, p. 57; RABELO *et al.*, 2012, p. 21), comunidades virtuais de aprendizagem (GOZZI, 2012, p. 3) e análise sociotécnica (DRUZIANI; KERN; CATAPAN, 2012, p. 194).

O aumento na quantidade de pesquisas abordando a GC tem sido observado por autores como Brewer e Brewer (2010, p. 330), Serenko *et al.* (2010, p. 3) e Singh e Soltani (2010, p. 145). Para Krogh, Nonaka e Rechsteiner (2012, p. 240), o estudo do conhecimento organizacional já tem se estabelecido com base em vários trabalhos teóricos e empíricos. Segundo Pastor, Santana e Sierra (2010, p. 2452), a GC já alcançou grande popularidade nas esferas acadêmicas e organizacionais nos últimos anos.

Assim sendo, tendo em vista a diversidade de estudos envolvendo a GC (não somente em assuntos distintos, como também em abordagens e métodos de pesquisa) e o estabelecimento do tema como uma área muito próxima à consolidação teórica (se já não totalmente consolidada), pergunta-se: quais os termos mais empregados em trabalhos publicados nos últimos dez anos? Com relação a este quesito, existem diferenças entre as publicações brasileiras e estrangeiras? O objetivo da presente pesquisa é, portanto, aplicar a mineração de textos, para extrair os termos mais empregados em resumos de pesquisas publicadas no período de 2003 a 2012, em bases de dados nacionais e internacionais.

O presente estudo encontra-se dividido nas seguintes seções: fundamentação teórica sobre GC e mineração de textos, procedimentos metodológicos adotados, apresentação e análise dos resultados, considerações finais e referências utilizadas.

2 Fundamentação teórica

A presente seção apresenta os conceitos fundamentais norteadores do estudo. Há que se esclarecer que não existe a pretensão de esgotar os assuntos, mas, sim, fornecer os subsídios mínimos para a compreensão da pesquisa e seu contexto.

2.1 Gestão do conhecimento

Antes da definição sobre o conceito de GC, faz-se necessária uma reflexão acerca do termo 'conhecimento'. Autores, como Sveiby (1998, p. 33) e Nonaka (1991, p. 162), concordam que não há um consenso, muito menos uma definição perfeita sobre o conceito de 'conhecimento', porém definem algumas características. Para Sveiby (1998, p. 35-42), (1) o conhecimento é tácito, difícil de expressar e dependente das experiências dos indivíduos; (2) é orientado para a ação; (3) é sustentado por regras, que estão atreladas ao resultado das ações; e (4) o conhecimento está em constante mutação. Já Nonaka (1991, p. 165) ressalta três características do conhecimento: (1) diz respeito a crenças e compromissos, ou seja, depende das experiências do indivíduo; (2) está relacionado à ação; e (3) é específico ao contexto e relacional. O autor citado sugere que a criação do conhecimento organizacional é decorrente de funções e responsabilidades gerenciais, desenho organizacional e práticas de negócios (NONAKA, 1991, p.164).

Uma vez constatada a existência de um contexto organizacional capaz de instigar a geração do conhecimento, naturalmente surgiram estudos que abordavam o processo de gestão deste contexto. Para Wiig (1997, p. 6), os anos entre 1986 e 1989 foram marcados por inúmeros relatos de pesquisas preocupados, explicitamente, em como gerenciar o conhecimento. Segundo Sveiby (2001, p. 5), o termo 'gestão do conhecimento' foi empregado, pela primeira vez, no contexto organizacional (existem registros anteriores de utilização deste termo, porém, em pesquisas relacionadas à inteligência artificial), justamente por Karl Wiig, que, segundo o autor citado, em 1990, publicou o primeiro artigo que continha estas palavras no título.

De acordo com Dalkir (2005), existem várias abordagens para explicar o ciclo da gestão do conhecimento nas organizações, sendo que, apesar disto, as mesmas apresentam mais similaridades entre si, do que disparidades. Basicamente tratam de descrever os seguintes processos (DALKIR, 2005, p. 43):

a) captura e criação: para Camelo-Ordaz, García-Cruz, Sousa-Ginel e Valle-Cabrera (2011, p.1444), o conhecimento dos indivíduos fornece a matéria-prima necessária para a criação do conhecimento organizacional. Cabe às empresas desenvolver condutas capazes de reter e transformar os conhecimentos individuais em coletivos (PROBST; RAUB; ROMHARDT, 2002, p. 87);

b) compartilhamento e disseminação: segundo Aktharsha, Anisa e Ali (2012, p. 23), estes processos são responsáveis por disseminar ou trocar conhecimento explícito e tácito, ideias, experiências, habilidades e tecnologias entre indivíduos ou grupos de funcionários, através da interação entre eles; e

c) aquisição e aplicação: Dalkir (2005, p. 70) associa a aplicação (ou a aquisição) sistemática do conhecimento em situações relevantes, diretamente à disposição da empresa em inovar - nominando esta ação de 'inteligência organizacional'.

O tema GC não está isento de críticas. Para Ibrahim e Reid (2010, p. 25), não se trata de um conceito reconhecido mundialmente, carecendo de definições comuns amplamente aceitas. Conforme Grant e Grant (2008, p. 578-579) a GC tem sido associada a um modismo na área de negócios (como tantos outros), além disto, destacam o fato de que a estrutura conceitual básica da GC nada mais é do que gestão da informação. Malhotra (2004, p. 90-91) é mais incisivo: trata-se de sistemas de informação. Já Pesqueux (2008, p. 6) observa que, na melhor das hipóteses, podem-se gerenciar apenas as condições para criar, formalizar e trocar conhecimento - mas nunca, o conhecimento em si.

2.2 Mineração de textos

Segundo Faro, Giordano e Spampinato (2012, p. 62), o objetivo primário da mineração de textos é encontrar conhecimento oculto em textos e apresentá-lo de forma coerente e concisa. Para Wang, Chang e Kao (2010, p. 494), trata-se de um conceito e estrutura propostos há mais de uma década atrás, desenvolvidos para extrair informação valiosa em grandes quantidades de documentos não estruturados ou semiestruturados. Já Yoon, Phaal e Probert (2008, p. 54) observam que as técnicas de mineração de textos (uma classe particular da mineração de dados) auxiliam a extração de conhecimento, a partir de dados textuais não estruturados. Em pesquisas recentes, esta abordagem tem sido empregada para estruturar o conteúdo de documentos em palavras-chave, identificando assuntos, resumindo seus conteúdos e visualizando as relações entre os documentos (YOON; PHAAL; PROBERT, 2008, p. 54).

Segundo Gajzer (2010, p. 223-224), o processo de mineração de textos é composto por quatro estágios:

a) transformação do texto: consiste na transformação do documento para um formato texto, removendo e/ou substituindo símbolos desnecessários;

b) separação de palavras (ou tokenização): para Soares, Prati e Monard (2008, p. 3), trata-se de um método que identifica as características importantes de um texto, a partir da quebra de fluxo contínuo de caracteres, removendo elementos textuais como pontuação, separação de sílabas, marcações e números (os quais, de maneira isolada, agregam pouco valor à informação);

c) *stemming*: segundo Soares, Prati e Monard (2008, p. 4), neste estágio, a quantidade de *tokens* é diminuída, a partir da extração dos sufixos e prefixos que formam cada *token*, ou seja, ocorre uma "normalização linguística, na qual as formas variantes de um termo são reduzidas a uma forma comum denominada *stem*"; e

d) matriz de frequências: conforme Fan *et al.* (2006, p.79), este estágio é responsável por categorizar os *stems* e associando-os às respectivas frequências de ocorrência nos textos analisados e, portanto, possibilitando inferências sobre suas proximidades, distâncias, sinônimos e termos relacionados.

A estes estágios, Soares, Prati e Monard (2008, p. 7) observam a importância de se estabelecer frequências para *tokens* unidos com outros adjacentes ou, até mesmo, mais distantes, visto que uma sequência de termos pode conter mais informações do que os termos isolados. Este procedimento é chamado de *n-grama*.

Com o estabelecimento da Internet como meio de disponibilização de informações, na forma de textos não estruturados, das mais variadas naturezas, é temeroso estabelecer os limites das aplicações da mineração de texto. Porém, Weiss, Indurkha e Zhang (2010, p. 189-194) apontam para aplicações que, se já não podem ser consideradas rotineiras, certamente estão próximas disto: sumarização, que consiste na criação de sumários a partir de vários textos do mesmo tópico; aprendizado ativo, que permite a classificação automática de documentos; e aprendizado a partir de dados não classificados.

3 Procedimentos metodológicos

A partir da classificação proposta por Silva e Menezes (2005, p. 19-22), a presente pesquisa pode ser enquadrada (1) quanto à sua natureza, como básica; (2) quanto à abordagem do problema, quantitativa; (3) do ponto de vista dos objetivos, exploratória; e (4) quanto aos procedimentos técnicos, como bibliográfica.

Os dados analisados foram obtidos de fontes primárias, a saber: bases de dados EBSCO, *Web of Knowledge*, SCOPUS, ScienceDirect e Scielo Brasil. Os dados (resumos) foram coletados a partir da busca por artigos que contivessem o termo 'Gestão do conhecimento' ou '*Knowledge management*' no título. Neste sentido, as bases internacionais trouxeram 5091 artigos (sendo que, destes, registraram-se 1337 duplicados e 297 sem o resumo, portanto, os registros de ambos os casos foram excluídos), contra 87 da base Scielo Brasil. Na tentativa de equilibrar a disparidade entre o número de registros, buscou-se, manualmente, artigos dos seguintes *sites* de eventos nacionais: Encontro de Administração da Informação, Encontro da ANPAD, Encontro Nacional de Engenharia de Produção, KMBRASIL e Simpósio de Administração da Produção, Logística e Operações Internacionais, bem como as revistas *Alcance*, *Brazilian Administration Review*, *Base*, *Brazilian Business Review*, *Biblos*, *Ciência da Informação*, *Data Grama Zero*, *Gestão & Produção*, *Informação & Informação*, *Informação & Sociedade*, *Perspectivas em Ciência da Informação*, *Revista de Administração Contemporânea*, *Revista de Administração de Empresas*, *Revista de Administração Mackenzie*, *Revista de Administração Pública*, *Revista de Administração da Universidade de São Paulo*, *Revista Brasileira de Biblioteconomia e Documentação*, *Revista Digital de Biblioteconomia e Ciência da Informação*, *Revista Eletrônica de Administração*, *Revista Portuguesa e Brasileira de Gestão e Transinformação*. Assim sendo, os artigos em português totalizaram 407 (deste número, 27 foram eliminados, visto que estavam presentes em mais de uma fonte).

É importante que se esclareça que (1) as bases pesquisadas foram escolhidas por conveniência, por estarem à disposição da instituição a qual dos autores da presente pesquisa estão lotados; (2) a busca pelos termos '*Knowledge management*' (nas bases internacionais) e '*Gestão do conhecimento*' (nas bases nacionais), no campo 'Título', é decorrente do fato de que algumas bases não permitem buscas por outros campos; (3)

estabeleceu-se um recorte temporal referente aos anos de 2003 a 2012, nas bases de dados – nas bases de revistas e eventos, foram recuperados artigos até o ano de 2011; e (4) o fato do estudo não ser estatístico, bem como as bases apresentarem nítida disparidade entre os números de artigos, os resultados não podem ser generalizados – devendo ser tomados, apenas, como indicativos.

De maneira geral, a pesquisa obedeceu ao protocolo estabelecido no Quadro 1:

Quadro 1 – Protocolo de pesquisa

Etapa	Objetivos	Materiais
1 Levantamento de artigos	Resgatar artigos que contenham o termo 'Gestão do conhecimento' e/ou 'Knowledge management' no título	Bases de dados, sites e anais de revistas e eventos
2 Seleção de artigos	Remoção de artigos duplicados (encontrados em mais de uma base de dados)	EndNote® X6 (THOMSON REUTERS, 2013)
3 Conversão de artigos	Separação e conversão dos artigos, do formato XML (gerado pelo EndNote® X6) para o formato txt	Aplicativo PHP desenvolvido pelo autor
4 Mineração de texto	Identificação e agrupamento de termos encontrados nos arquivos convertidos	PreText 2 (LABORATORY OF COMPUTATIONAL INTELLIGENCE, 2013)
5 Pré-análise de resultados	Identificação e retirada de stems naturalmente recorrentes encontrados nos resumos (exemplo: 'objetivo', 'pesquisa', 'study', 'paper', ...)	

Fonte: Dados da pesquisa.

Ressalta-se que, na Etapa 4, houve a necessidade de ajustar trechos do código fonte do aplicativo PreText 2, visto que registrou-se a ocorrência, no estágio de *stemming*, de alguns *tokens* com o mesmo radical, por formarem *stems* distintos – situação esta já prevista por autores, como Rijsbergen (1979, p. 4) e Martins *et al.* (2003, p. 5). Além disto, nesta mesma etapa, parametrizou-se o *software* PreText 2 (LABORATORY OF COMPUTATIONAL INTELLIGENCE, 2013), de forma a retornar *stems* que estivessem presentes em pelo menos cinco resumos, com o intuito de reduzir a diversidade de *stems* encontrados. As análises foram limitadas aos 20 primeiros *stems* que mais ocorreram.

Finalmente, há que se esclarecer o fato de que, por se tratar de uma pesquisa exploratória, existe reduzida (ou nenhuma) margem para a explicação dos resultados à luz de estudos da mesma natureza. Ainda assim, na medida em que as evidências encontradas forem sendo apresentadas, as mesmas serão submetidas a uma reflexão empírica, de maneira a prover elementos necessários à elaboração de hipóteses a serem testadas em pesquisas posteriores.

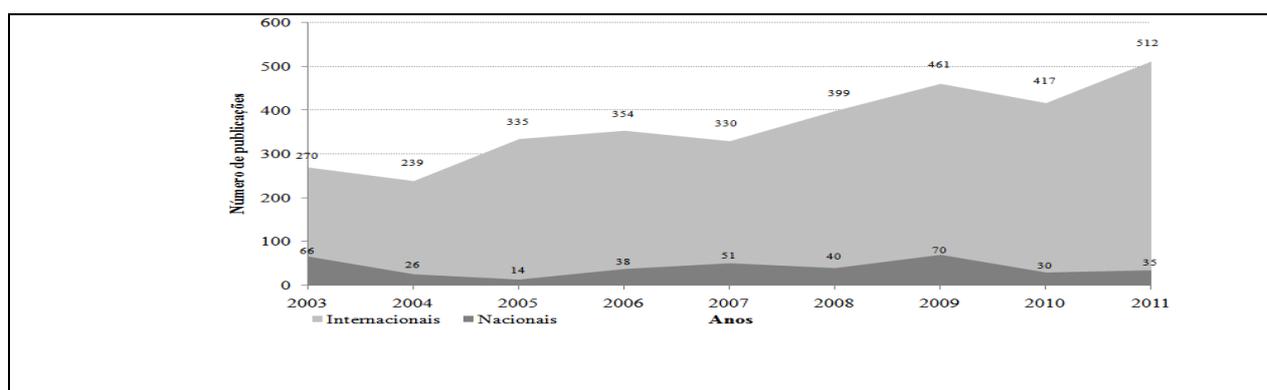
4 Apresentação e análise dos resultados

Inicialmente, há que se apresentar a quantidade de artigos, com os termos 'Gestão do conhecimento' e 'Knowledge management', no título, distribuídos ao longo dos anos.

A

Figura 1 mostra estes dados:

Figura 1 - Publicações de GC de 2003 a 2011



Fonte: Dados da pesquisa.

Desconsiderando-se a diferença (óbvia) entre a quantidade de publicações entre as bases nacionais e internacionais, nota-se que, enquanto as bases de dados nacionais apresentam certa estabilidade na quantidade de artigos resgatados entre os anos de 2003 a 2011 (girando em torno da média de 41,11, com desvio padrão de 18,33), as publicações encontradas nas bases internacionais mostram crescimento neste período, apesar da presença de ciclos de baixa a cada três anos – vale informar que se excluiu o ano de 2012 da comparação, visto que não se buscou artigos em *sites* de revistas e eventos nacionais (ainda assim, na base Scielo Brasil, encontraram-se 10 artigos com o termo 'Gestão do conhecimento' no título e 436 nas bases internacionais).

Em relação aos *stems* 1-grama encontrados nas bases nacionais e internacionais, formaram-se, respectivamente, 3353 e 10412. A

Tabela 1 apresenta aqueles que atingiram as 20 primeiras posições:

Tabela 1 – 1-grama

Nacionais				Internacionais			
<i>stems</i>	f_i	n	%	<i>stems</i>	f_i	n	%
organizac	792	267	70,26	organizat	4191	1503	43,48
empres	495	196	51,58	system	3206	1383	40,01
process	441	215	56,58	process	3152	1535	44,40
informac	406	184	48,42	develop	2730	1549	44,81
desenvolv	289	173	45,53	inform	2607	1315	38,04
tecnolog	284	144	37,89	model	2364	1108	32,05

pratic	248	125	32,89	practic	2213	1137	32,89
model	227	115	30,26	technolog	1864	1065	30,81
estrateg	208	123	32,37	design	1668	962	27,83
competit	172	110	28,95	effect	1625	1010	29,22
sistem	156	86	22,63	perform	1564	715	20,68
ambient	148	99	26,05	implement	1547	942	27,25
ferrament	144	89	23,42	busi	1455	810	23,43
relacion	128	94	24,74	learn	1379	625	18,08
criaca	124	93	24,47	support	1359	926	26,79
identific	124	97	25,53	firm	1277	536	15,50
compartilh	112	79	20,79	strategi	1188	574	16,60
inov	105	56	14,74	compani	1151	646	18,69
estrutur	99	79	20,79	innov	1122	512	14,81
projet	96	54	14,21	framework	1116	636	18,40

Fonte: Dados da pesquisa.

A

Tabela 1 foi composta daqueles *stems* que, de alguma forma, representam temas ou abordagens à GC – excluíram-se dela *stems*, que, naturalmente, são encontrados em resumos, como 'pesq', 'studi', 'research', além daqueles que obviamente estão presentes em artigos que tratam de GC, como 'conhec', 'gc', 'manag', 'knowledg'. A maioria dos *stems* (13) apresentados na

Tabela 1 é equivalente entre si, nos artigos em bases nacionais e internacionais, não somente no significado, como, também, no percentual de artigos em que aparecem. Além disto, as posições em que os *stems* se localizam em cada grupo também não se distanciam de forma diametralmente opostas, mostrando certo alinhamento com que estes termos são empregados, independentemente das bases de dados pesquisadas. Já naqueles que não apresentam correspondência nas 20 primeiras posições, em bases nacionais, envolvem termos associados a alguns processos de GC ('criação', 'identificação', 'compartilhamento'), enquanto que, nas bases internacionais, destacam-se aqueles relacionados, de alguma forma, à medição ('effectiveness', 'performance').

Para os *stems* 2-grama (formados pela ocorrência de dois termos adjacentes), o *software* PreText 2 (LABORATORY OF COMPUTATIONAL INTELLIGENCE, 2013), da maneira como parametrizado, formou 19774 de língua portuguesa e 111136 *stems* de língua inglesa.

A Tabela 2 mostra os 20 primeiros: Tabela 2 – 2-grama

Nacionais				Internacionais			
<i>stems</i>	f_i	n	%	<i>stems</i>	f_i	n	%
conhec_organizac	112	78	20,53	manag_system	893	582	16,84
technolog_informac	90	65	17,11	knowledg_share	501	317	9,17
pratic_gesta	73	50	13,16	manag_knowledg	446	364	10,53
process_gesta	68	51	13,42	competit_advantag	367	295	8,53
vantag_competit	60	48	12,63	inform_tecnolog	340	274	7,93
compartilh_conhec	46	39	10,26	practic_implic	313	299	8,65
conhec_tacit	42	26	6,84	manag_process	284	195	5,64
model_gesta	41	38	10,00	manag_practic	275	155	4,48
informac_conhec	39	31	8,16	decis_make	274	198	5,73

conhec_empres	35	30	7,89	inform_system	265	194	5,61
capital_intelectual	34	28	7,37	implement_knowledg	257	218	6,31
sistem_gesta	31	24	6,32	knowledg_creation	243	186	5,38
criaca_conhec	30	25	6,58	organizat_knowledg	241	188	5,44
estrateg_organizac	23	22	5,79	km_process	229	118	3,41
ferrament_tecnolog	23	19	5,00	tacit_knowledg	215	148	4,28
aprendizag_organizac	23	16	4,21	km_practic	214	134	3,88
gesta_informac	22	16	4,21	km_strategi	211	117	3,38
cultur_organizac	22	15	3,95	human_resourc	208	138	3,99
recurs_human	21	20	5,26	organizat_perform	197	95	2,75
ambient_organizac	20	18	4,74	effect_knowledg	195	167	4,83

Fonte: Dados da pesquisa.

Dos 20 *stems* formados a partir dos resumos das bases nacionais e internacionais, 11 deles são correspondentes. Os demais, aparentemente, não chegam a formar um padrão, ainda que, dentre os nacionais, destacam-se aqueles relacionados a termos mais abstratos, como 'ambiente' e 'cultura' organizacionais – entre os internacionais, novamente surgem termos associados à medição ('*organizational performance*' e '*effectiveness knowledge*'), além de práticas de GC ('*practices implications/implicit*' e '*km practices*'). Além disto, é curioso o fato de que o termo 'estratégia' estar associado à 'organização' nos artigos nacionais e, nos internacionais, à própria GC.

A

Tabela 3 apresenta os resultados para os 20 primeiros *stems* 3-grama (de um total de 21667 formados em artigos localizados em bases nacionais e 151734 em bases internacionais):

Tabela 3 – 3-grama

Nacionais				Internacionais			
<i>stems</i>	f_i	n	%	<i>Stems</i>	f_i	n	%
pratic_gesta_conhec	59	39	10,26	knowledg_manag_system	772	506	14,64
process_gesta_conhec	57	41	10,79	knowledg_manag_process	249	168	4,86
gesta_conhec_organizac	49	39	10,26	implement_knowledg_manag	240	203	5,87
model_gesta_conhec	27	24	6,32	knowledg_manag_practic	228	124	3,59
gesta_conhec_empres	22	21	5,53	knowledg_manag_strategi	156	94	2,72
sistem_gesta_conhec	20	16	4,21	effect_knowledg_manag	152	128	3,70
conceit_gesta_conhec	15	14	3,68	knowledg_manag_organizat	113	96	2,78
relacion_gesta_conhec	14	12	3,16	relationship_knowledg_manag	105	87	2,52
tecnolog_informac_comunicaca	13	12	3,16	develop_knowledg_manag	89	86	2,49
ferrament_tecnolog_informac	12	10	2,63	small_medium_size	89	76	2,20
ferrament_gesta_conhec	11	11	2,89	inform_commun_tecnolog	87	85	2,46
conhec_tacit_explicit	11	11	2,89	critic_success_factor	83	65	1,88
vantag_competit_sustent	10	8	2,11	success_knowledg_manag	82	68	1,97
program_gesta_conhec	10	8	2,11	structur_equat_model	80	72	2,08
estrateg_gesta_conhec	9	6	1,58	knowledg_manag_tool	76	56	1,62
informac_gesta_conhec	9	8	2,11	sustain_competit_advantag	76	67	1,94
instrument_gesta_conhec	9	7	1,84	knowledg_manag_activ	76	55	1,59
process_criaca_conhec	8	7	1,84	knowledg_manag_implement	75	52	1,50
relaca_gesta_conhec	8	7	1,84	approach_knowledg_manag	75	70	2,02
cientif_gesta_conhec	7	7	1,84	knowledg_manag_initi	75	58	1,68

Fonte: elaborado pelos autores.

Fonte: Dados da pesquisa.

Dos 10 *stems* que não apresentam correspondência entre as bases de dados, destacam-se, nas bases nacionais, aqueles que podem remeter a trabalhos acadêmicos ('conceito gestão do conhecimento' e 'científico gestão do conhecimento'), bem como à percepção de 'instrumentos'/'ferramentas' associadas à GC ou, especificamente, TICs. Nas bases internacionais, termos relacionados à implementação/desenvolvimento da GC ('*implementation knowledge management*', '*development knowledge management*' e '*knowledge management implementation*') são recorrentes, além daqueles que remetem ao sucesso – provavelmente da implantação – ('*critical success factors*' e '*successful knowledge management*') ou às atividades/iniciativas de GC ('*knowledge management initiatives*' e '*knowledge management activities*').

A partir da análise 3-grama, os resultados que trazem sequências de *stems* já formam frases que pouco acrescentam às particularidades esperadas. Por exemplo, na análise 4-grama, nas bases nacionais, encontraram-se apenas seis resultados: 'gesta_conhec_poder_ser' (10 ocorrências em 10 artigos), 'importanc_gesta_conhec_organizac' (7/7), 'gesta_conhec_ter_ser' (6/6), 'trabalh_ter_objet_apresent' (6/6), 'artig_apresent_result_pesquis' (5/5) e, 'dad_ser_colet_mei' (5/5).

De maneira geral, surpreende o fato do *stem* 'inov' (relacionado às palavras associadas à 'inovação'), em artigos em português, ser encontrado apenas na 17ª posição e o *stem* 'innov', nos artigos em língua inglesa, na 19ª posição, nos *stems* 1-grama e não ter sido encontrado nos demais, visto ser este termo recorrentemente associado à GC por autores como Aktharsha, Anisa e Ali (2012, p.22); Krogh, Nonaka e Rechsteiner (2012, p.258); Camelo-Ordaz, García-Cruz, Sousa-Ginel e Valle-Cabrera (2011, p.1442).

4 Considerações finais

Inicialmente, há que se considerar as perguntas que originaram a pesquisa e como elas foram respondidas. Neste sentido, os termos mais empregados em trabalhos publicados, nos últimos dez anos, foram obtidos a partir do processo de mineração de textos e expressos nas tabelas 1, 2 e 3. Quanto à segunda pergunta ('existem diferenças entre as publicações brasileiras e estrangeiras?'), os resultados obtidos demonstram que, na medida em que seja possível realizar inferências a partir de 20 palavras isoladas (mesmo com elevadas frequências), os artigos encontrados em bases internacionais apresentam termos mais associados aos aspectos pragmáticos da GC. Porém, registra-se que os resultados evidenciam alinhamento em relação a determinados temas – o que pode sugerir certo amadurecimento conceitual.

Ainda assim, o estudo apresenta certas limitações. Duas delas são técnicas e arbitrarias: o número mínimo de cinco artigos para que os *stems* fossem recuperados e a quantidade de 20 *stems* analisados. Porém,

trata-se de parâmetros que podem ser alterados de acordo com os objetivos da pesquisa. No presente estudo, observa-se uma relação inversamente proporcional entre o aumento do número mínimo de artigos em que os *stems* encontram-se e a diversidade de *stems* analisados, sendo que, neste sentido, há que se encontrar um meio termo capaz de atender aos anseios do pesquisador. Outra limitação deve ser comentada. O fato de se vislumbrar determinada tendência nos artigos avaliados, a partir de termos isolados (ou mesmo adjacentes), é arriscado – o contexto todo, no qual o termo se encontra, certamente fornece mais subsídios. Todavia, é uma questão de *trade-off*: Poucas (ou, talvez, nenhuma) técnicas permitiriam a análise contextual simultânea da quantidade de resumos avaliados (380 nacionais e 3457 de bases internacionais). Contudo, como a pesquisa é exploratória (e, portanto, os resultados servem mais como *insights* para trabalhos futuros), julga-se que a forma como as evidências foram apresentadas não compromete a significância do estudo.

Os resultados encontrados remetem para questões envolvendo tanto a GC, como a mineração de textos em artigos científicos. Quanto ao primeiro aspecto, as evidências apresentadas foram obtidas com base nos *stems* localizados nas 20 primeiras posições – o aumento de *stems* analisados confirmariam os resultados encontrados? Ressalta-se que para chegar aos 20 primeiros *stems* para as análises 1, 2 e 3-grama em português, retiraram-se respectivamente 27, 17 e 16 *stems* (nas bases internacionais eliminaram-se 16, 34 e 22 *stems*). Além disto, é possível que os termos mais empregados, de forma geral, mudem no transcorrer dos anos? Quanto ao segundo aspecto, ainda que não estritamente associado com o tema GC em si, mas, sim, com a técnica de mineração de textos empregada, o transcorrer do presente trabalho trouxe um questionamento (e, portanto, sugestão para pesquisas futuras): as palavras-chave dos artigos, e até mesmo os títulos, apresentam relação direta com os termos empregados nos resumos?

De maneira geral, é possível afirmar que os termos empregados nos artigos publicados em bases nacionais e internacionais de 2003 a 2012, com 'Gestão do conhecimento' no título, aparentam mais alinhamento do que distinções. Termos como 'práticas', 'processos', 'tecnologias de informação', 'sistemas', parecem apontar para a consolidação da GC nestes aspectos. Os resultados, porém, chamam mais a atenção pelas dissimilaridades. Enquanto que as bases nacionais mostram termos recorrentemente associados aos elementos mais abstratos da GC, nas internacionais registra-se a predominância de palavras que indicam pesquisas mais voltadas a questões pragmáticas. Neste sentido, a mineração de textos pode auxiliar na determinação de possíveis tendências de pesquisas na área de GC, contribuindo, assim, para o estabelecimento de estudos mais alinhados ao contexto mais vigente.

Referências

- AKTHARSHA, U. S.; ANISA, H.; ALI, S. D. The usage behavior of knowledge management system in hospitals. *The IUP Journal of Knowledge Management*, v. 10, n. 2, p. 22-44, 2012.
- BREWER, P. D.; BREWER, K. L. Knowledge management, human resource management, and higher education: a theoretical model. *Journal of Education for Business*, v. 85, n. 6, p. 330-335, 2010.
- CAMELO-ORDAZ, C. *et al.* The influence of human resource management on knowledge sharing and innovation in Spain: the mediating role of affective commitment. *The International Journal of Human Resource Management*, v. 22, n. 7, p. 1442-1463, Apr. 2011.
- CAMPOS, L. F. B. Análise da nova gestão do conhecimento: perspectivas para abordagens críticas. *Perspectivas em Ciência da Informação*, v. 12, n. 1, p.104-122, jan./abr. 2007.
- CAPUANO, E. A.; CASAES, J.; COSTA, J. R. Inteligência competitiva e suas conexões epistemológicas com gestão da informação e do conhecimento. *Ciência da Informação*, v. 38, n. 2, p. 19-34, maio/ago. 2009.
- DALKIR, K. *Knowledge management in theory and practice*. Oxford: Elsevier, 2005.
- De NOMI, I.; ORSI, L.; PILOTTI, L. Uncertainty and forecasting of future extreme events in organizations: new approaches in knowledge management. *The Icfai University Journal of Knowledge Management*, v. 7, n. 3-4, p. 96-113, 2009.
- DRUZIANI, C. F. M.; KERN, V. M.; CATAPAN, A. H. A gestão e a engenharia do conhecimento aliadas na modelagem do conhecimento – análise sistêmica CISM e contextual Commonkads de um repositório na Web. *Perspectivas em Gestão & Conhecimento*, v. 2, n. 1, p. 194-217, jan./jun. 2012.
- FAN, W. *et al.* Tapping the power of text mining. *Communications of the ACM*, v. 49, n. 9, p. 77-82, Sept. 2006.
- FARO, A.; GIORDANO, D.; SPAMPINATO, C. Combining literature text mining with microarray data: advances for system biology modeling. *Briefings in Bioinformatics*, v. 13, n. 1, p. 61-82, jan. 2012.
- GAJZER, M. Text and data mining techniques in aspect of knowledge acquisition for decision support system in construction industry. *Technological and Economic Development of Economy*, v. 16, n. 2, p. 219-232, 2010.
- GONZALEZ, R. V. D.; MARTINS, M. F.; TOLEDO, J. C. Gestão do conhecimento em uma estrutura organizacional em rede. *Perspectivas em Ciência da Informação*, v. 38, n. 1, p. 57-73, jan./abr. 2009.

GOZZI, M. P. O processo de gestão do conhecimento em comunidades virtuais de aprendizagem. *Perspectivas em Gestão & Conhecimento*, v. 2, n. 2, p. 3-14, jul./dez. 2012.

GRANT, K. A.; GRANT, C. T. Developing a model of next generation knowledge management. *Issues in Informing Science and Information Technology*, v. 5, p. 571-590, 2008.

IBRAHIM, F.; REID, V. Unpacking knowledge management: management fad or real business practice? *Enterprise Risk Management*, v. 2, n. 1, p. 24-38, 2010.

KROGH, G.; NONAKA, I.; RECHSTEINER, L. Leadership in organizational knowledge creation: a review and framework. *Journal of Management Studies*, v. 49, n. 1, p. 240-277, 2011.

LABORATORY OF COMPUTATIONAL INTELLIGENCE. *PreText 2: a reestruturação da ferramenta de pré-processamento de textos*. São Carlos: LABIC/ICMC/USP, 2013. Disponível em <<http://sites.labic.icmc.usp.br/pretext2>>. Acesso em: 7 jun. 2013.

LEE, L. T. S.; SUKOCO, B. M. The effects of entrepreneurial orientation and knowledge management capability on organizational effectiveness in Taiwan: the moderating role of social capital. *International Journal of Management*, v. 24, n. 3, p. 549-572, Sept. 2007.

LEITE, F. C. L.; COSTA, S. M. S. Gestão do conhecimento científico: proposta de um modelo conceitual com base em processos de comunicação científica. *Ciência da Informação*, v. 36, n. 1, p. 92-107, jan./abr. 2007.

LUNDVAL, B. A.; NIELSEN, P. Knowledge management and innovation performance. *International Journal of Manpower*, v. 28, n. 3/4, p. 207-223, 2007.

MALHOTRA, Y. Why knowledge management systems fail? Enablers and constraints of knowledge management in human enterprises. In: KOENIG, M. E. D.; SRIKANTIAH, T. K. *Knowledge management lessons learned: what works and what doesn't*. [s.l.]: Information Today Inc, 2004. , p. 87-112

MARTINS, C. A. *et al. Uma experiência em mineração de textos utilizando clustering probabilístico clustering hierárquico*. São Carlos: Instituto de Ciências Matemáticas e de Computação, julho 2003. Relatório técnico.

NONAKA, I. The knowledge-creating company. *Harvard Business Review*, v. 69, n. 6, p.160-171, jul.aug. 1991.

PASTOR, I. M. P.; SANTANA, M. P.; SIERRA, C. M. Managing knowledge through human resource practices: empirical examination on the Spanish automotive industry. *The International Journal of Human Resource Management*, v. 21, n. 13, p. 2452-2467, Oct. 2010.

PERROT, B. E. A strategic risk approach to knowledge management. *Business Horizons*, v. 50, n. 6, p. 523-533, Nov./Dec. 2007.

PESQUEUX, Y. Une critique du knowledge management. *EuroEconomica*, v. 20, n. 1, p. 5-16, 2008.

PROBST, G.; RAUB, S.; ROMHARDT, K. *Gestão do conhecimento*. Porto Alegre: Bookman, 2002.

RABELO, R. A. et al. Gestão do conhecimento em processos de transformação organizacional: o desenvolvimento da intimidade como fator facilitador. *Perspectivas em Gestão & Conhecimento*, v. 2, n. 1, p. 21-35, jan/jul 2012.

RIJSBERGEN, C. J. Automatic text analysis. In: RIJSBERGEN, C. J. *Information Retrieval*. Newton: Butterworths, 1979. Disponível em <<http://www.dcs.gla.ac.uk/Keith/Chapter.2/Ch.2.html>>. Acesso em: 7 jun. 2013.

SERENKO, A. et al. A scientometric analysis of knowledge management and intellectual capital academic literature (1994-2008). *Journal of Knowledge Management*, v. 14, n. 1, p. 3-23, 2010.

SHANG, S.; LIN, S. F. Understanding the effectiveness of Capability Maturity Model Integration by examining the knowledge management of software development processes. *Total Quality Management*, v. 20, n. 5, p. 509-521, May 2009.

SILVA, E. L.; MENEZES, E. M. *Metodologia da pesquisa e elaboração de dissertação*. 4. ed. Florianópolis: UFSC, 2005.

SINGH, A.; SOLTANI, E. Knowledge management practices in Indian information technology companies. *Total Quality Management*, v. 21, n. 2, p.145-157, Feb. 2010.

SOARES, M.; PRATI, R. C.; MONARD, M. C. PreText: a reestruturação da ferramenta de pré-processamento de textos. São Carlos: ICMC, ago. 2008. Relatório técnico.

SVEIBY, K. E. *A nova riqueza das organizações: gerenciando e avaliando patrimônios de conhecimento*. Rio de Janeiro: Campus, 1998.

SVEIBY, K. E. *Knowledge management: Lessons from the pioneers*. 2001. Disponível em: <www.sveiby.com/Portals/0/articles/KM-lessons.doc>. Acesso em: 23 maio 2013.

THOMSON REUTERS. *EndNote*[®]. Carlsbad: Thomson Reuters, 2013. Disponível em <<http://endnote.com>>. Acesso em: 7 jun. 2013.

WANG, M. Y.; CHANG, D. S.; KAO, C. H. Identifying technology trends for R&D planning using TRIZ and text mining. *R&D Management*, v. 40, n. 5, p. 491-509, 2010.

WEISS, S. M.; INDURKHYA, N.; ZHANG, T. *Fundamentals of predictive text mining*. London: Springer-Verlag, 2010.

WIIG, K. M. Knowledge management: an introduction and perspective. *The Journal of Knowledge Management*, v. 1, n. 1, p. 6-14, Sept. 1997.

YOON, B.; PHAAL, R.; PROBERT, D. Morphology analysis for technology roadmapping: application of text mining. *R&D Management*, v. 38, n. 1, p. 51-68, 2008.

ZANG, S.; GAO, Q.; YANG, C. A new method for design process knowledge management. *Journal of Advanced Manufacturing Systems*, v. 7, n. 1, p. 107-110, 2008.