

## PROPOSTA DE METODOLOGIA PARA A RECUPERAÇÃO DA PRODUÇÃO CIENTÍFICA EM CIÊNCIA DA INFORMAÇÃO NA BASE BRAPCI

### RESUMO

Propõe uma metodologia de atribuição de pesos a campos de busca em processos de Recuperação da Informação (RI) na Base Referencial de Artigos de Periódicos em Ciência da Informação (Brapci). Define características de uma metodologia de RI para determinar a relevância de estudos em uma área temática específica. Realiza um teste piloto na Brapci para a validação da proposta, com o tema epistemologia da Ciência da Informação. Discute duas vertentes, a primeira, relativa à recuperação da informação e a segunda, sobre a temática de busca para ilustrar o alcance do método. A proposta de hierarquização de artigos por pesos, atribuídos aos campos título, palavra-chave e resumo, categoriza os documentos mais relevantes para a temática buscada. Conclui que os resultados da aplicação do método adequam-se às bases teóricas da área e à importância dos seus autores para o domínio específico e que a produção recuperada abrange documentos relevantes para outras áreas relacionadas à CI, reiterando seu caráter interdisciplinar.

**Palavras-chave:** Recuperação de Informação. Método de Recuperação de Informação. Bases de dados. Epistemologia da Ciência da Informação.

### PROPOSITION OF A METHODOLOGY FOR THE RETRIEVAL OF SCIENTIFIC PRODUCTION IN INFORMATION SCIENCE ON BRAPCI DATABASE

#### ABSTRACT

This work proposes the construction of a methodology destined to assign weights to different search fields, in order to improve Information Retrieval (IR) from Brazilian Database for Periodicals on the Domain of Information Science (BRAPCI). It questions the characteristics of an IR methodology used to determine the relevance of studies in a specific thematic field. It performs a pilot test in Brapci, in order to validate the proposed method, using the theme "Information Science epistemology". It discusses a theoretical reference turned to two branches; the first, related to information retrieval, and the second, whose character is analytical, taking this search theme in order to illustrate the range of this method: the Information Science epistemology. The proposition of hierarchizing articles by their weight, in the search fields "title", "keywords", and "abstract", has the aim of categorize the most relevant documents in the search. It concludes that the results obtained by the application of this method are adequate both to the theoretical framework of the field and to their authors' importance in the specific domain. This assumption reiterates the interdisciplinary character of this domain, which is already recognized in literature.

**Keywords:** Information Retrieval. Information Retrieval Method. Database. Information Science Epistemology

**Juliana Lazzarotto  
Freitas**

Mestranda do Curso de Pós-Graduação em Ciência, Gestão e Tecnologia da Informação da Universidade Federal do Paraná - UFPR. Bolsista Capes.

[julianalf@ufpr.br](mailto:julianalf@ufpr.br)

**Leilah Santiago Bufrem**

Doutora em Ciências da Comunicação pela Universidade de São Paulo - USP. Pós-Doutora pela Universidad Autonoma de Madrid - UAM. Professora Titular do Departamento de Ciência e Gestão da Informação da Universidade Federal do Paraná – UFPR.

[leilah@ufpr.br](mailto:leilah@ufpr.br)

**Rene Faustino Gabriel  
Junior**

Mestrando do Curso de Pós-Graduação em Ciência, Gestão e Tecnologia da Informação da Universidade Federal do Paraná - UFPR. Bolsista Capes.

[rene.gabriel@ufpr.br](mailto:rene.gabriel@ufpr.br)

## 1 INTRODUÇÃO

As características de um contexto de crescimento contínuo de informações disponíveis, em que o usuário depara-se com dificuldades para obter resultados relevantes e precisos em suas buscas por informação, evidenciam a validade de estudos sobre as estruturas dos instrumentos para a representação e recuperação da informação, almejando soluções ou métodos adequados de Recuperação da Informação (RI).

O estudo tem como escopo a proposta de um método de RI para a Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (Brapci) baseado no grau de importância atribuído aos seus campos textuais de busca. A partir da atribuição de pesos ou pontos a esses campos e suas combinações, procurou-se contribuir com a recuperação e a relevância do processo de análise de conteúdo nos repositórios de conhecimento científico. A Brapci teve seu modelo operacionalizado na *Web* e constitui-se no mais completo repositório da produção científica periódica da área de Ciência da Informação (CI) no Brasil.

O estudo piloto, que se opta por denominar de método ou metodologia de RI ao longo do texto, baseia-se no pressuposto teórico de que os campos de busca, desde os estudos de Anderson (1971) até a atualidade, têm sido determinados como títulos, palavras-chaves, resumos e introdução, pois se constituem nos campos mais ricos para se obterem informações preliminares sobre os documentos.

A possibilidade de identificação da pertinência relativa à busca de um documento acadêmico-científico pode trazer contribuições significativas aos pesquisadores em suas estratégias de investigação. Essa contribuição se efetiva, tanto na economia de tempo na tentativa de encontrar informações pertinentes em meio a uma infinidade de documentos, como também na facilidade para realizar análises do conteúdo dessa literatura, simplificando os processos complexos e manuais de identificação e obtenção desses documentos.

A partir da identificação de conceitos relativos à temática selecionada, foram estabelecidos descritores de busca. A criação um método baseado na atribuição de pontos aos elementos permitiu a hierarquização dos artigos recuperados, assim como a análise da relevância dos resultados obtidos.

A escolha do tema “delimitação epistemológica do campo da Ciência da Informação” remete à preocupação em se analisar o domínio dessa ciência, a fim de contribuir com sua consolidação. Para validar o método aqui sugerido de RI, foi moldado um relatório que contivesse a somatória dos pontos para facilitar a visualização das respostas de busca, correspondendo ao teste piloto aplicado a uma área selecionada dentro da Brapci.

Deve-se considerar que a dificuldade em encontrar literatura pertinente que tratasse especificamente da atribuição de pesos a campos de busca e sua combinação para a hierarquização de documentos recuperados em bases de dados incentivou, de maneira especial, a realização desta proposta de recuperação. Isso porque se pode percebê-la como um impulso para o desenvolvimento de outros estudos.

Dentre as outras motivações para a realização da pesquisa, relativas à escolha do tema para o teste de validação do método, está a preocupação em estimular as análises da produção científica da área de CI e de facilitar a compreensão deste domínio do conhecimento.

Como argumenta Bufrem (2008, p. 38), a publicação científica periódica

tornou-se, em seu processo histórico, um instrumento indispensável tanto como meio de promoção acadêmica, quanto como modalidade de promoção e fortalecimento do ciclo de criação, organização e difusão do conhecimento.

Assim, leva-se em conta para este estudo a contribuição social e científica desse tipo de fonte de informação. Ao facilitar a compreensão desse ciclo de conhecimento, a publicação periódica também tem sido valorizada para a compreensão do domínio científico da CI, que vem sendo discutido, especialmente tendo-se em vista a chamada interdisciplinaridade como fator de complexidade para sua delimitação.

Logo, há necessidade de compreender o domínio científico desse campo de conhecimento específico, procurando-se na literatura sobre sua formação, especialmente aquela voltada à epistemologia, com vistas à reconstrução de sua base teórico-prática. Essa reconstrução é proposta com base na análise de domínio, introduzida na CI por Hjørland e Albrechtsen (1995) que enfatizam a especialidade, a disciplina, o domínio e o ambiente como unidades de estudo. Por meio da análise de domínio, a informação é estudada em contextos específicos do conhecimento, ou seja, em contextos distintos e por isso é tida

como objeto subjetivo, sujeito a diferentes definições e interpretações, dependendo do ambiente em que está inserida.

O domínio analisado para a validação do método de recuperação da informação aqui proposto refere-se às questões epistemológicas relacionadas à área de CI. Deve-se ter em conta que o estudo da epistemologia de uma área é determinante para a construção dos métodos científicos desta área, a metodologia por sua vez é fator essencial para a construção do conhecimento científico. Segundo Rendón-Rojas (2008), as exigências metodológicas de cada área se estabelecem a partir dos pressupostos epistemológicos dos quais se parte.

As metodologias podem ser variadas e empregar enfoques pragmatistas, construtivistas, fenomenológicos, sociais, hermenêuticos e em determinados momentos é possível empregar métodos formais como a lógica e a comprovação estatística e empírica (RENDÓN-ROJAS, 2008).

A organização epistemológica de uma área refere-se à base teórica conceitual que delimita determinado campo de estudo, ou seja, o embasamento teórico de uma ciência já constituída.

## **2 TRAJETÓRIA METODOLÓGICA**

Este estudo, de natureza exploratório-metodológica, propõe um método de RI em uma base de dados da área da CI, validado por meio de um teste piloto de busca sobre um tema particular. A pesquisa exploratória foi amparada pela análise de um quadro teórico sobre CI e RI, não exaustivo, porém suficiente para fundamentar a metodologia proposta.

Para a criação e aplicação do método de atribuição de pesos aos campos de busca a fim de facilitar o processo de recuperação de registros na base Brapci e sua conseqüente análise, realizou-se uma investigação na literatura da área da CI, que se desdobra em duas vertentes, a primeira, relativa à RI e a segunda, de caráter analítico, decorre da temática de busca escolhida para ilustrar o alcance do método,. Já, na segunda fase da pesquisa, a busca por esses conceitos possibilitou a identificação de descritores associados ao tema escolhido: epistemologia, teoria do conhecimento, teoria da ciência, filosofia da ciência, ciência da

ciência, história da ciência, gnosiologia e termos que remetessem à epistemologia como: epistemológico, epistemológica, epistemológicos e epistemológicas.

A partir da definição dos descritores, na terceira fase foi realizada uma busca na base Brapci, que contempla a produção científica de trinta periódicos nacionais da área da CI, do ano de 1972 a 2009. Com a aplicação dos descritores anteriormente definidos nos campos título, palavra-chave e resumo foram recuperados 84 documentos, que apresentaram ao menos um dos descritores estabelecidos em algum dos campos. A tipologia de documentos delimitou-se a ensaios, pontos de vista, artigos científicos, comunicações, revisões de literatura e relatos de casos.

A quarta etapa consistiu na construção do modelo de pesos para permitir a hierarquização dos documentos. Foi utilizada a contagem binária de três *bits*, sendo o primeiro menos significativo (com peso um) atribuído ao resumo, o segundo *bit* (com peso dois) atribuído às palavras-chave e o terceiro mais significativo (com peso quatro) atribuído ao título. A representação binária possibilita a variação de oito números indo de zero a sete, conforme atribuição relacionada ao Quadro 1.

Bit	Atribuição	Valor decimal
1	Resumo	1
2	Palavras-chave	2
3	Título	4

**Quadro 1** – Atribuição dos valores aos campos de busca  
Fonte: Os autores (2010)

O modelo binário permite somatória de *bits* do menos para o mais significativo, possibilitando obter uma pontuação mais relevante quando os descritores estiverem inseridos no título, seguido da palavra-chave e menor no resumo. Essa combinação permite que quando o descritor estiver no resumo e na palavra-chave (somando três pontos) não ultrapasse a pontuação de quando estiver somente no título (quatro pontos, mais relevante).

No caso do descritor estar inserido em todos os elementos, a somatória obterá a pontuação máxima, correspondente a sete pontos. E, no menos expressivo dos casos, quando o descritor aparecer somente no resumo, o registro obterá um ponto. Logo, a

hierarquização desse caso com menor expressividade (um ponto) nunca será superior à dos documentos que contiverem os descritores somente na palavra-chave, com dois pontos, ou somente no título, com quatro pontos.

Entre as possibilidades podem-se obter combinações, conforme Quadro 2.

Pontos	Descritor
0 ponto	Nenhum dos elementos
1 ponto	Dentro do resumo
2 pontos	Dentro das palavras-chave
3 pontos	Dentro das palavras-chave e resumo
4 pontos	Dentro do título
5 pontos	Dentro do título e resumo
6 pontos	Dentro do título e palavras-chave
7 pontos	Dentro do título, palavras-chave e resumo

**Quadro 2** – Pontos atribuídos aos documentos pelas combinações de pesos  
Fonte: Os autores (2010)

Considerados esses critérios, a totalização dos pontos foi feita individualmente para cada artigo recuperado da Brapci, por meio de uma ferramenta integrada, com a finalidade de avaliar o peso relativo de cada artigo. Dos processos resultaram: total de pontos; pontos do título; pontos das palavras-chave; pontos do resumo; a referência da publicação; o nome da revista; o volume; o número; o ano da publicação e o resumo.

Os dados obtidos foram exportados como tabela para o *Microsoft Excel*<sup>®</sup>, o que possibilitou uma hierarquização dos textos de acordo com a pontuação recebida. Com os dados coletados foi feita uma tabulação, com a separação de autores e a representação por meio de gráficos para os que mais produziram sobre o tema, para os autores que obtiveram maior peso, e também para as palavras-chave de maior incidência, utilizadas como uma ferramenta de apoio para a análise dos conceitos tratados pelos autores.

Como última etapa, realizou-se a análise qualitativa dos resultados da busca, a fim de se comprovar se a realidade da pesquisa no domínio em questão foi representada integralmente pelo método proposto. Esta análise consistiu na leitura dos 21 documentos, a fim de se observar a relevância do assunto e sua conformidade com os objetivos do tema proposto. Foi, também, realizada uma leitura para confirmar a possibilidade da baixa relevância de possíveis artigos, visto que não foi descartada a possibilidade da presença de documentos cujo conteúdo não fosse revelado explicitamente.

### 3 RECUPERAÇÃO DA INFORMAÇÃO EM BASES DE DADOS

A Recuperação em larga escala torna-se mais precisa e eficiente quando feita em bases e bancos de dados, instrumentos de organização e recuperação das informações, grandemente alicerçados pela informática. Segundo Lopes (2002, p. 65), os bancos de dados, entendidos como sistemas de RI, apresentam uma complexidade inerente às estruturas que os suportam, no processo de armazenamento e busca de informação, envolvendo uma série de aspectos interdependentes. Um destes aspectos, de acordo com a autora, é o planejamento de estratégias de busca específicas, a fim de se obter qualidade na informação recuperada. Esta qualidade exige cuidados e atenção a indicadores, conforme alerta Bufrem (2008, p. 2):

A análise sobre indicadores de qualidade é parte das atividades de um processo em que se clarificam e distinguem a provisão da base e o seu monitoramento. Entretanto, considerar que os indicadores não são um fim em si mesmo, mas instrumentos práticos para a administração de bases de dados e tomadas de decisão implica o encaminhamento dado ao seu processo de assunção, encaminhamento que, esse sim, volta-se aos propósitos ou fins para os quais se construiu o objeto.

Assim, destacam-se os indicadores, conforme a consistência interna da base de dados, por um lado e, por outro, em relação ao usuário e ao contexto em que se situam as relações de busca e recuperação.

Concorda-se com Lancaster (1998), que ressalta a qualidade da estratégia de busca e o vocabulário como fatores importantes para a atividade de busca, entretanto, não se pode esquecer a problemática que envolve a qualidade das bases de dados, hoje fontes de informação consideradas imprescindíveis. Assim, a forma de representação e descrição do conteúdo dos textos atribuída pelos autores ou pela revista foi sempre priorizada. Logo, projetou-se uma organização hierárquica de documentos, ponderando essas representações e descrições pré-estabelecidas, especialmente pelas facilidades de indexação advindas das tecnologias emergentes.

Ao referir-se à marcante abordagem tecnológica relativa ao conceito de RI até o momento em que se começa a ter uma preocupação maior com o usuário, Martins (2008, p. 86) adota uma visão histórica recuperando a influência das ciências cognitivas sobre a CI,

que dariam lugar às aproximações posteriores, de caráter social. Este caráter tem sido considerado tão importante quanto o tecnológico, pois visa o compartilhamento do conhecimento através da realização das análises de domínio. Entre os autores que assim argumentam pode-se considerar Shera, Wersig e Capurro (*apud* MARTINS, 2008, p. 86), ao afirmarem que os sistemas deixam de cumprir seu papel principal quando passam a centrar-se muito mais na sofisticação tecnológica do que nas necessidades dos usuários e suas interações sociais.

Essa visão epistemológica de cunho social reflete uma evolução das concepções da CI, pela qual se recuperam os elementos fundantes e delimitadores de uma dada área do conhecimento, por meio de análises de domínio. Essa compreensão afeta os modos de organização e recuperação da informação, pautados, por sua vez, nas dimensões física, cognitiva e social da área, podendo-se afirmar que os sistemas de organização do conhecimento e seus processos são compreendidos a partir dos estudos de domínio.

Alguns domínios têm alto grau de consenso e critérios de relevância explícitos, mas outros como a CI podem apresentar aspectos conflitantes que, se por um lado dificultam a estruturação e consolidação de uma área, por outro, trazem novas possibilidades de inovação nas formas de produção, organização, comunicação e disseminação da informação.

#### **4 ATRIBUIÇÃO DE PESOS A CAMPOS DE BUSCA**

Para Arif, Rahman e Mukta (2009), a atribuição de pesos para descritores de busca é a idéia central para qualquer técnica de recuperação informacional que se pauta na hierarquização de documentos. Os pesos dos termos para a recuperação informacional são estimados a partir da medida de co-ocorrência de descritores do documento. Estes são localizados conforme seus graus em uma hierarquia baseada em algoritmos representativos da importância relativa de cada termo em gráficos, relacionando-os à estratégia de busca do usuário.

Em um estudo de Strzalkowski (1999), relatou-se a construção de um método de busca para a RI em linguagem natural e sua eficácia na representatividade da recuperação de documentos inteiros. Os autores revelam que a *Natural Language Information Retrieval*



permite uma melhor representação do conteúdo de documentos, em especial quando se definem termos compostos e grupos de palavras para a busca, frases expressivas que podem denotar conceitos importantes para o domínio da base de dados.

Para os autores, uma busca é transformada em uma coleção ponderada de termos, que podem ser advindos diretamente do texto ou indiretamente, através de sinônimos ou mapas de domínio. O objetivo, segundo eles, é obter os documentos sobre o mesmo tema, colocados próximos, e os de diferentes temas colocados o mais distante possível. Entretanto nem sempre as fórmulas de ponderação baseadas em termos de distribuição são ideais, apesar de que, conforme os autores, o processo de atribuição de pesos a termos usando combinações foi o método mais eficaz visualizado nos resultados.

Sobre os três principais componentes dos algoritmos utilizados para a operacionalização de sistemas de RI, de acordo com Mittendorfer e Winiwarter (2001), o primeiro deles é utilizado para a análise das estruturas das buscas, outro, para o processamento das palavras de busca e o último para a combinação dos dois anteriores em um resultado final.

Assim sendo, um sistema de RI que visa aprimorar o desempenho da recuperação informacional também pode ser balizado em um esquema de atribuição de pesos aos termos de documentos, auxiliando na identificação dos termos que melhor representam o conteúdo dos documentos, de acordo com Camacho-Guerrero e Macedo (2005, p. 171). Esses pesos também podem ser conferidos aos principais campos de busca das bases, já que os termos inseridos nos campos de busca específicos: títulos, palavras-chave e resumos, seguramente são de maior representatividade para a identificação temática dos documentos.

Para Kobashi e Fernandes (2009, p. 674), “a convergência tecnológica modifica as condições de produção, acesso e recepção de bens simbólicos”, o que justifica segundo as autoras, a reflexão epistemológica como condição para pensar os novos modos de responder aos problemas relativos à informação.

## 5 TEMÁTICA RECUPERADA NA BRAPCI: EPISTEMOLOGIA DA CIÊNCIA DA INFORMAÇÃO

As relações entre a epistemologia e a CI destacam-se sempre que se pretende analisar e definir seus domínios e objetos de estudo, estabelecer as relações interdisciplinares e transdisciplinares da constelação de áreas que se integram para a construção de conhecimentos e analisar os princípios metodológicos que orientam suas atividades de pesquisa.

Este estudo é orientado especialmente pelas noções relativas à epistemologia, suas funções e definições e aos pontos de vista de autores da área sobre a delimitação epistemológica da CI.

A importância da investigação epistemológica na CI, de acordo com Rendón-Rojas (2008), se dá por três motivos. O primeiro deles é para que ocorra um autoconhecimento da disciplina, para que seja realizada uma busca pela identidade da CI. O segundo é para que o corpo teórico desse domínio do conhecimento possa se fundamentar, criando um corpo conceitual preciso, claro e definido. E o terceiro motivo é a interdisciplina, ou seja, interagir com outras disciplinas, mas com uma identidade bem definida, a fim de não invadir outros campos do conhecimento e não deixar ser absorvida por outras áreas.

A importância de estudos sobre a construção histórica de conceitos científicos se justifica a partir da situação já observada, que a CI emprestou conceitos de outras disciplinas, muitas vezes “sem uma reflexão epistemológica do impacto das possíveis incongruências e incompatibilidades teóricas para a delimitação de sua identidade científico-disciplinar” (RABELLO 2008, p. 30).

As relações entre campos tem sido preocupação de autores que tentam conciliar seus estudos históricos com as tendências que se configuram. Desse modo, Freire (2008, p. 3) propõe a perspectiva de Ginzburg para analisar os indícios da constituição de um espaço de pesquisa sobre a Epistemologia da CI nesse domínio. O paradigma indiciário é uma competência cognitiva originária do tempo em que a sobrevivência da nossa espécie dependia da observação para obter o conhecimento sobre os animais a serem caçados, seus hábitos e trilhas nas correntes migratórias.

Entretanto, autores privilegiam aspectos que, sob seu ponto de vista, contribuem mais especificamente para a constituição teórico-metodológica dessa ciência. É o caso de

Barreto (2008), para quem a CI se constrói ao sabor das inovações na tecnologia. Para melhor compreender esse processo, o autor considera didático e fundamental contar a história de como se atuava no passado e como se verificou a evolução das práticas da área, de modo a contribuir também para a formação de seus profissionais.

Com essa intenção, Araújo (2006, p. 3) elabora uma reflexão epistemológica sobre o “fazer científico” estruturado na ciência moderna em termos teóricos e metodológicos e nas tecnologias da informação, em termos aplicados, visto que, essa configuração é que possibilitou o surgimento da CI. Contudo, considera que a mesma não garantirá o pleno desenvolvimento da área como campo do conhecimento consistente e moderno e que a ciência moderna enquanto visão e prática científica encontra-se esgotada e as tecnologias de informação são apenas mecanismos auto-regulados que funcionam segundo princípios de automatismos. Portanto, com a finalidade de apontar bases teóricas e metodológicas para uma epistemologia da CI, fundamentada nos pressupostos e paradigmas de Bachelard (1996) e Ginzburg (1991) citados por Araújo (2006, p. 7), propõe respectivamente, o conceito de ciência formativa, a qual parte dos três estados do espírito científico. Apóia-se nas condições psicológicas do progresso científico e no conceito de paradigma indiciário, cuja proposta é a intuição empírica e racional como base metodológica para “o fazer científico”. É necessário que as disciplinas possuam identidade própria para poderem interatuar entre si, sem invadir ou absorver o espaço da outra.

A reflexão epistemológica sobre a ciência da informação é necessária para seu ulterior desenvolvimento e fortalecimento, tanto disciplinar interno como social externo. (RENDÓN-ROJAS, 2008).

A reflexão sobre o caráter social da CI apresenta-se em estudos que avaliam concepções de pesquisadores. É o caso do estudo de Araújo et al (2007) em que apresenta os resultados de uma pesquisa realizada com professores de todas as faculdades brasileiras da área de Biblioteconomia e CI, na qual se buscou apresentar argumentações dos docentes sobre o caráter social, interdisciplinar e pós-moderno da CI e também visualizar a natureza de sua relação com a Biblioteconomia. Ainda como parte da pesquisa, foi solicitado a cada professor que indicasse dez autores e dez obras que julgasse relevantes na área de CI. Houve uma grande dispersão nas indicações. Os autores mais citados foram: Saracevic, Lancaster,

Borko e Pinheiro, e a obra mais citada foi “Ciência da informação: origem, evolução e relações”, de Saracevic.

É neste contexto que se pretende visualizar a delimitação epistemológica do campo da CI, pois os aspectos históricos e sociais que permearam a criação desta ciência, junto à necessidade de se ter um discurso científico que trate especificamente dos processos de produção à disseminação da informação são suficientemente relevantes para o campo ser consolidado e respeitado no meio científico. Considera-se que o próprio objeto deste domínio, a informação, é responsável pela sua interdisciplinaridade, pois é o elemento tido como precursor da criação de novos conhecimentos e da comunicação destes em todos os campos do saber.

## **6 RESULTADOS**

Os resultados da busca permitiram a identificação de um total de 84 documentos, dos quais dezessete obtiveram a pontuação máxima de sete pontos, apresentando os descritores estabelecidos nos três campos de busca: título, palavra-chave e resumo. Com seis pontos, foram levantados quatro documentos, cujos descritores foram encontrados em dois campos, no título e na palavra-chave; com a pontuação cinco identificaram-se cinco documentos, nos campos de título e resumo; já com pontuação quatro foram identificados três documentos com a presença dos descritores somente no campo título; e com três pontos recuperaram-se quatorze documentos nos campos palavra-chave e resumo. A inexpressiva quantidade de documentos com a pontuação máxima, isso é, apenas dezessete artigos representados por descritores nos três campos, demonstra que poderia ser aperfeiçoada a atribuição dos elementos representativos do conteúdo.

Por outro lado, 40 documentos retrataram o termo em mais de um campo, ou seja, esse número de documentos obteve possibilidade de combinação por apresentar um dos descritores em mais de um campo de busca. Não foram identificados documentos com descritores somente nas palavras-chave, o que demonstra coerência na representação.

Por meio da hierarquização dos pontos foram selecionados os 21 documentos que obtiveram maior peso, especificamente os que somaram sete e seis pontos. A partir desses

artigos, realizou-se uma análise de conteúdo com o objetivo de comprovar se a metodologia de pontuação refletiu o conteúdo dos artigos recuperados.

Houve 41 documentos cujos descritores estabelecidos foram encontrados somente no resumo, com pontuação inexpressiva para a metodologia proposta. Diante desse resultado, foi realizada uma leitura dos resumos dos artigos sem pontuação relevante para a metodologia de pesos, a fim de identificar se acaso o método descartou documentos pertinentes à temática. A análise realizada a partir da leitura demonstrou ser apenas aparente essa inexpressividade.

A Tabela 1 representa o resultado da recuperação conforme a categorização realizada.

**Tabela 1** – Resultado da recuperação

<b>Critério</b>	<b>Total de documentos</b>
Título, palavras-chave e resumo (7 pontos)	17
Título e palavras-chave (6 pontos)	4
Título e resumo (5 pontos)	5
Título (4 pontos)	3
Palavras-chave e resumo (3 pontos)	14
Somente resumo (1 ponto)	41
<b>Total</b>	<b>84</b>

Fonte: Os autores (2010)

Na Recuperação foram identificados os autores mais produtivos na temática, não necessariamente os mais representativos, apresentados na Tabela 2.

**Tabela 2** – Autores que mais publicaram sobre a temática

<b>NOME DO AUTOR</b>	<b>NÚMERO DE ARTIGOS</b>
GONZÁLEZ DE GÓMEZ	4
FRANCELIN	3
MOSTAFA	3
CAMPOS	2
CAPUANO	2
ESTABEL	2
GARCÍA GUTIÉRREZ	2
MEDEIROS	2
MORO	2
RENDÓN-RÓJAS	2
SANTAROSA	2
VENÂNCIO	2

Fonte: Os autores (2010)

## 7 ANÁLISE

Os dados recuperados e categorizados, conforme o método descrito, permitiram a realização da análise de conteúdo dos documentos.

Observa-se que os autores Francelin (2003 e 2005), Freire (2008), González de Gómez (2001), Medeiros e Marques (2003), Oddone (2007), Rabello (2008), Rendón-Rojas (2008) e Zandonade (1999-2000), responsáveis pela autoria dos 17 documentos com o grau de hierarquização mais elevado, representados por sete pontos, foram os que fundamentaram teoricamente a pesquisa realizada previamente à recuperação dos documentos da Brapci e cujos textos foram analisados em seu conteúdo integral. Isso significa dizer que a metodologia proposta revelou uma adequação com as bases teóricas da área e com a importância dos seus autores para o domínio específico.

Por outro lado, Araújo (2006) e Barreto (2008), embora obtendo pontuação cinco, estão entre os mais referenciados na revisão de literatura sobre delimitação epistemológica da CI, prioritariamente a autores que apresentaram pontuação máxima de sete pontos, como Saldanha (2008), Shera (1977) e Zoboli, Silva e Bordas (2006). Isso se deve ao fato de que esses últimos trataram mais especificamente sobre a epistemologia, em suas perspectivas social e crítica e não diretamente relacionada com as questões da CI.

Com peso sete na hierarquização, um texto de Roach (2008), denominado “Abordagem fenomenológico-hermenêutica e pesquisa em educação: um estudo de vigilância epistemológica”, tratou da epistemologia e teoria do conhecimento relacionada ao domínio da pesquisa educacional, sem relação, portanto, com a área tratada nesta pesquisa, a CI. Esse fato mostra que a base Brapci, repositório da produção científica desse domínio, devido à sua característica interdisciplinar, traz pesquisas de relevância também para outras áreas do conhecimento.

Essas condições sugerem que as revistas da área de CI são receptivas à produção de áreas interdisciplinares, as quais contribuem com a formação e desenvolvimento do domínio em pauta. Entretanto, isso não fragiliza a proposta metodológica deste estudo, pois ajuda a destacar essa relação entre áreas.

Outra informação obtida foi que nem todos os pesquisadores com maior quantidade de publicações sobre a temática foram os que apresentaram mais alto índice

nessa hierarquia. Destacam-se entre os primeiros González de Gómez, com quatro documentos, Francelin com três, Medeiros com dois e Rendón-Rojas também com dois. Estes se destacaram tanto pela expressividade de seu trabalho para a área temática da epistemologia da CI como no aspecto quantitativo, obtendo maior representatividade de publicações comparados aos outros autores, que em sua maioria apresentaram somente um documento relevante.

Por meio da análise dos autores mais produtivos, aponta-se Capuano, um dos que se destacaram na análise quantitativa com dois documentos, e também na qualitativa por apresentar a atribuição de cinco pontos na hierarquia. O autor trata, em um de seus estudos, sobre as relações epistemológicas entre a estrutura secular das organizações no setor público brasileiro e os conceitos de democracia e cidadania, e propõe um modelo genérico de organização fundamentada na informação e no conhecimento. Em seu segundo texto, fala sobre as conexões epistemológicas entre a inteligência competitiva e a gestão da informação e do conhecimento. Esse autor, de acordo com os resultados do modelo de RI proposto, tem pouca representatividade para a realização do estudo da delimitação epistemológica da CI, já que é aplicado a estudos da Gestão da Informação no meio organizacional.

Já, os que se destacaram somente no aspecto quantitativo, ambos com duas publicações e sem presença alguma nos níveis de hierarquia pertinentes, ou seja, que não apresentaram os descritores estabelecidos da área temática em mais de um campo, foram: Estabel, Moro e Santarosa (2009) e Campos e Venâncio (2006 e 2007).

García-Gutiérrez (2006) ainda teve seu nome citado na hierarquia de três pesos, que se refere à obtenção de descritores simultaneamente em palavras-chave e resumo. Contudo, seu texto aborda a análise do conhecimento e de sua organização, a partir da chamada epistemografia interativa, que se opõe à epistemologia. A epistemografia visa acolher o conhecimento excluído, incorporando a ele e à sua organização as questões éticas, culturais e políticas. O que entra em desacordo com os estudos de delimitação das ciências, facilitados pelo estudo da epistemologia.

Mostafa (1983 e 2003), com o total de três publicações sobre a temática, tem uma delas com valor de três pontos. Neste artigo, ela faz uma discussão comparativa das três correntes epistemológicas presentes no espaço de interrelação entre a comunicação e a

educação: o humanismo, o criticismo e o pós-estruturalismo. Mostafa aponta o olhar pós-estruturalista em direção aos catálogos, os resumos, os indicadores e as listas de citações. Logo, esta autora não entra em questões relativas à CI.

Na outra obra da mesma autora, com pontuação um, encontra-se uma análise sobre a comunicação educativa para visualizar tendências epistemológicas na produção científica dessa área. E em sua terceira obra, também com peso um, a autora analisa a produção na área da Biblioteconomia, com referência à teoria do conhecimento dentro da relação sujeito/objeto.

Dos 21 documentos que obtiveram a pontuação mais relevante entre seis e sete pontos, todos tiveram seu conteúdo condizente ao tema proposto na pesquisa, com variações gerais, específicas e associativas de enfoques, porém pertinentes à proposta do método de RI planejado, abrangendo a quase totalidade dos autores que discutem sobre a epistemologia.

Com pontuação cinco, ou seja, a presença de descritores em resumo e título foram recuperados cinco documentos e somente um, por Pereira Junior, com o título de *“Glutamatergic Mechanisms of Perceptual Learning: an Essay in Neuroepistemology”* (2006) não se adequou à temática.

Na outra etapa de análise, que consistiu na leitura dos resumos de 41 documentos, os quais apresentaram os descritores somente no campo de resumo (um ponto) foi verificada a existência de cinco documentos representativos para a temática e que não entraram nas categorias de maior pontuação, pois estavam dispersos em meio aos considerados menos expressivos pelo método.

Esta situação é passível de ocorrer devido a três motivos, o primeiro deles é quando a representação dos documentos realizada pelos seus autores não conseguiu refletir adequadamente seu conteúdo nos campos propostos para a análise e indicados pela literatura como determinantes para a identificação do conteúdo de documentos. A segunda razão é que não foi descartada a possibilidade da presença de algum texto que propositalmente não explicitasse pelo autor, o seu conteúdo. A terceira possibilidade é do conteúdo da epistemologia estar embasando estudos teórico-práticos de outros domínios do conhecimento.



No primeiro caso identificam-se os artigos intitulados “Perspectivas em informação: tendências e tensões entre abordagens físicas, cognitivas e emergentes”, de Campos e Venâncio, (2007) e também o artigo “Paradigmas contemporâneos da Ciência da Informação: a recuperação da informação como ponto focal”, de Almeida et al (2007), o qual utiliza o estudo da epistemologia para facilitar a compreensão sobre a complementaridade das dimensões física, cognitiva e social que influenciam a área da RI, porém não usa o termo epistemologia nos campos que melhor identificam o documento.

No segundo caso, o artigo de Campos e Venâncio (2006), denominado “O objeto de estudo da ciência da informação: a morte do indivíduo”, não permite identificar seu real conteúdo por meio da representatividade do título, por usar uma linguagem figurativa.

E no último caso, o artigo “A formação de professores e a capacitação de bibliotecários com limitação visual por meio da EAD em ambiente virtual de aprendizagem” por Estabel, Moro e Santarosa (2009), aborda a epistemologia *Vygotskyana* como base do processo pedagógico para o processo de formação mencionado. Significa que os autores aplicam conhecimentos da epistemologia a outras áreas.

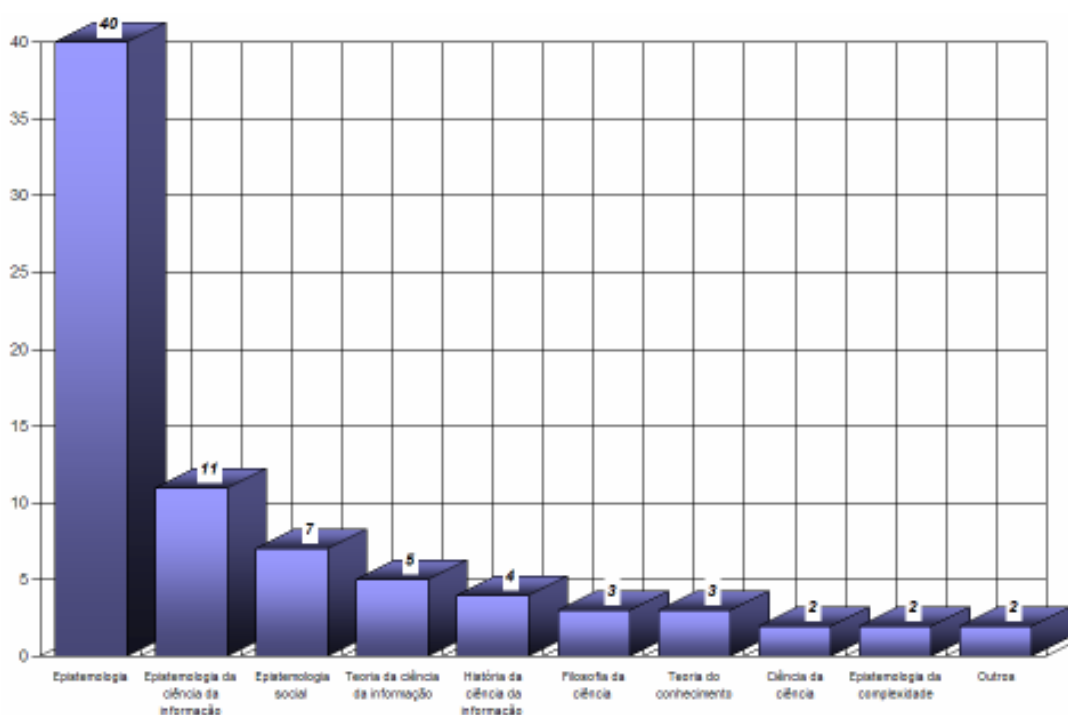
Uma pesquisa que na metodologia também foi classificada com baixa pontuação, e que ao final das análises foi considerada relevante para a área investigada, foi o trabalho de Pinheiro (2005), intitulado “Processo evolutivo e tendências contemporâneas da CI”. Essa dispersão ocorreu devido a uma distorção na metodologia de recuperação proposta, provocada pela indisponibilidade das palavras-chave e do texto completo pela revista, fazendo com que o artigo não fosse pontuado pelo sistema.

Como elemento complementar para a análise de conteúdo, optou-se por selecionar as palavras-chave mais incidentes nos resultados da recuperação, a fim de comprovar se a temática do estudo condiz com o conteúdo dos documentos.

Com o descritor epistemologia foram identificados 40 documentos, considerando-se os termos compostos, como por exemplo: epistemologia política, epistemologia genética, epistemologia construtivista, epistemologia da educação, epistemologia social e epistemologia interativa.

Notou-se, por meio da recuperação resultante da metodologia proposta, que o descritor gnosiologia não é mais utilizado para tratar da epistemologia, visto que, não foi encontrado em nenhum dos campos dos artigos recuperados da Brapci.

Quanto ao termo teoria do conhecimento, teve uma baixa incidência nas palavras-chave dos artigos, aparecendo somente em três documentos. Percebeu-se que houve uma diferenciação entre o termo epistemologia e teoria do conhecimento pelos autores, pois o conceito de epistemologia está mais relacionado ao estudo da ciência em si e à delimitação das ciências, já que cada uma delas apresenta sua própria epistemologia e a teoria do conhecimento parece estar mais focada nos estudos sobre o conhecimento humano, sua possibilidade e sua amplitude.



**Gráfico 1-** palavras-chave mais incidentes  
Fonte: Os autores (2010)

Observa-se no gráfico que as palavras-chave mais incidentes nos documentos são epistemologia, epistemologia da CI e epistemologia social. Isso reafirma que o modelo proposto tem sua área de concentração no embasamento teórico utilizado.

## 8 CONSIDERAÇÕES FINAIS

Como resultado concreto do estudo pode-se afirmar que o método amparou o processo de recuperação e análise dos documentos, dando maior visibilidade à identificação de temas, autores e relações entre eles.

Foram analisados conceitos relativos à temática em questão, estabelecendo-se descritores a fim de contribuir para a montagem da estratégia de busca. O método criado com base na atribuição de pontos ou pesos aos elementos permitiu a identificação e a hierarquização dos artigos recuperados de acordo com o planejamento. Além disso, foi analisada qualitativamente a relevância dos resultados quantitativos obtidos, comprovando-se a validade da metodologia. Entretanto, observam-se discrepâncias nos resultados, não decorrentes do método, mas sim da representação inadequada dos documentos.

A discrepância entre os valores e a representatividade temática decorre de três fatores. O primeiro refere-se à representação inadequada dos documentos pelos seus autores. O segundo, da possibilidade do autor não revelar propositalmente o conteúdo de seu texto. O terceiro provém do fato de o conteúdo da epistemologia estar embasando estudos teórico-práticos de outros domínios do conhecimento.

Este último fator explica o fato de na análise de categorias terem sido identificados documentos com alta pontuação, porém não relevantes à área de CI. Logo, retrata-se o caráter interdisciplinar da área, embora ainda persistam interpretações de estudiosos que não visualizam esse fenômeno como um aspecto positivo da CI.

Neste caso concreto, o método atendeu aos fins de RI sobre o tema Epistemologia da CI, permitindo uma maior facilidade na identificação e categorização desse conteúdo. Vale ressaltar que o estudo epistemológico serviu como alicerce para o desenvolvimento da metodologia de pesos em campos de busca aqui proposta.

Os documentos que obtiveram a pontuação mais relevante, entre seis e sete pontos, tiveram seu conteúdo condizente com o tema proposto na pesquisa, com variações gerais, específicas e associativas de enfoques, porém pertinentes à proposta da metodologia de RI planejada, compreendendo a quase totalidade dos autores que discutem sobre a epistemologia.

Assim sendo, nesta pesquisa exploratório-metodológica prevaleceu o intuito de incentivar estudos e análises da produção científica de domínios específicos do conhecimento, em especial da área da CI, o que significa que os resultados encontrados aqui não podem ser vistos como conclusivos e que estudos complementares abordando novas temáticas devem ser realizados para a consolidação do modelo.

## REFERÊNCIAS

ALMEIDA, D. P. R. et al. Paradigmas Contemporâneos da Ciência da Informação: a recuperação da informação como ponto focal. **Revista Eletrônica Informação e Cognição**, São Paulo, v. 6, n. 1, p. 16-27, 2007.

ANDERSON, M. D. **Book indexing**. Cambridge: *Cambridge University Press*, 1971.

ARAÚJO, C. A. Á. et al. A Ciência da Informação na visão dos professores e pesquisadores brasileiros. **Informação & Sociedade: Estudos**, João Pessoa. 2007.

ARAÚJO, E. A.. Por uma ciência formativa e indiciária: proposta epistemológica para a ciência da informação. **Encontros Bibli**, Florianópolis, v. 11, n. esp., p. 1-14, 2006.

ARIF, A. S. M.; RAHMAN, M. M.; MUKTA, S. Y. *Information retrieval by modified term weighting method using random walk model with query term position ranking*. In: INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING SYSTEMS, 15-17 May, Singapore, **Proceedings...** Singapore, 2009.

BARRETO, A. A. Uma quase história da ciência da informação. **DataGramaZero**, Rio de Janeiro, v. 9, n. 2, abr. 2008.

BUFREM, L. S. Práticas de organização e divulgação da produção intelectual em Ciência da Informação no Brasil. **Encontros Bibli**, Florianópolis, n. esp., p. 36-53, 1 sem 2008.

CAMACHO-GUERRERO, J. A.; MACEDO, A. A. Tecnologias de recuperação de informações na web. In: TEIXEIRA, C. A. C. (Org.). **Web e multimídia: desafios e soluções**, 2005. p. 167-186.

CAMPOS, L. F. B.; VENÂNCIO, L. S. Perspectivas em (in)formação: tendências e tensões entre abordagens físicas, cognitivistas e emergentes. **Transinformação**, Campinas, v. 19, n. 2, p. 107-118, maio/ago. 2007.

\_\_\_\_\_. O objeto de estudo da ciência da informação: a morte do indivíduo. **Informação & Informação**, Londrina, v. 11, n. 1, p. 1-21, jan./jun. 2006.

CAPUANO, E. A. et al. Inteligência competitiva e suas conexões epistemológicas com gestão da informação e do conhecimento. **Ciência da Informação**, Brasília, v. 38, n. 2, p. 19-34, maio/ago. 2009.

ESTABEL, L. B.; MORO, E. L. S.; SANTAROSA, L. M. C. A formação de professores e a capacitação de bibliotecários com limitação visual por meio da EAD em ambiente virtual de aprendizagem. **Transinformação**, Campinas, v. 21, n. 1, p. 23-32, jan./abr. 2009.

FRANCELIN, M. M. Abordagens em epistemologia: Bachelard, Morin e a epistemologia da complexidade. **Transinformação**, Campinas, v. 17, n. 2, p. 101-109, 2005.

\_\_\_\_\_. A epistemologia da complexidade e a ciência da informação. **Ciência da Informação**, Brasília, v. 32; n. 2, p. 64-68, 2003.

FREIRE, I. M. Um olhar sobre a produção científica brasileira na temática epistemologia da ciência da informação. **Tendências da Pesquisa Brasileira em Ciência da Informação**, João Pessoa, v. 1, n. 1, p. 1-31, 2008.

GARCÍA-GUTIÉRREZ, A. Cientificamente favelados: uma visão crítica do conhecimento a partir da epistemografia. **Transinformação**, Campinas, v. 18, n. 2, maio/ago. 2006.

GONZÁLEZ DE GÓMEZ, M. N. Para uma reflexão epistemológica acerca da Ciência da Informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 6, n. 1, jan./jun. 2001.

HJØRLAND, B; ALBRECHTSEN, H. *Toward a new horizon in information science: domain-analysis*. **Journal of the American Society for Information Science**, New York, v. 46, n. 6, p. 400-425, 1995.

KOBASHI, N. Y.; FERNANDES, J. C. Pragmática linguística e organização da informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 5., 2009, João Pessoa. **A responsabilidade social da Ciência da Informação**. João Pessoa: UFPB, 2009. p. 676-673.

LANCASTER, F. W. **Indexing and abstracting in theory and practice**. 2. ed. London: Library Association, 1998.

LOPES, I. L. Estratégia de busca na recuperação da informação. **Ciência da Informação**, Brasília, v. 31, n. 2, p. 60-71, mai./ago. 2002.

MARTINS, R. Informação e conhecimento: uma abordagem dos sistemas de recuperação de informações a partir das interações sociais. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 13, n. 2, p. 77-87, 2008.

MEDEIROS, A. M. S. de; MARQUES, M. A. R. B. Habermas e a teoria do conhecimento. **Educação Temática Digital**, Campinas, v. 5, n. 1, p. 1-24, dez. 2003.

MITTENDORFER, M.; WINIWARTER, W.: *Experiments with the Use of Syntactic Analysis in Information Retrieval: Proceedings of NLDB'01*, Madrid, Espanha, 2001. Disponível em: <http://citeseer.ist.psu.edu/mittendorfer01experiments.html>. Acesso em 3 dezembro de 2009.

MOSTAFA, S. P.; MÁXIMO, L. F. A produção científica da Anped e da Intercom no GT da Educação e Comunicação. **Ciência da Informação**, Brasília, v. 32, n. 1, jan./abr. 2003.

MOSTAFA, S. P. A produção de conhecimento em biblioteconomia. **Revista de Biblioteconomia de Brasília**, Brasília v. 11, n. 2, jul./dez. 1983.

\_\_\_\_\_. Catálogos, dispositivo de interpelação?. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 8, n. 2, jul./dez. 2003

ODDONE, N. Revisitando a “epistemologia social”: esboço de uma ecologia sociotécnica do trabalho intelectual. **Ciência da Informação**, Brasília, v. 36, n. 1, p. 108-123, jan./abr. 2007.

PEREIRA JUNIOR, A. *Glutamatergic Mechanisms of Perceptual Learning: an Essay in Neuroepistemology*. **Revista Eletrônica Informação e Cognição**, Marília, v. 5, n. 1, 2006.

PINHEIRO, L. V. R. Processo evolutivo e tendências contemporâneas da ciência da informação. **Informação & Sociedade: Estudos**, João Pessoa, v. 15, n. 1, p. 13-48. 2005.

RABELLO, R. et al. História dos conceitos e ciência da informação: apontamentos teórico-metodológicos para uma perspectiva epistemológica. **Encontros Bibli**, Florianópolis, n. 26, p. 17-46, 2º sem 2008.

RENDÓN-ROJAS, M. Á. La ciencia de la información en el contexto de las ciencias sociales y humanas. Ontología, epistemología, metodología e interdisciplina **DataGramZero**, Rio de Janeiro, v.9, n. 4, ago, 2008.

ROACH, E. F. F. Abordagem fenomenológico-hermenêutica e pesquisa em educação: um estudo de vigilância epistemológica. **Educação Temática Digital**, Campinas, v. 10, n. 1, p. 198-226, dez. 2008.

SALDANHA, G. S. Thomas Kuhn na epistemologia da Ciência da Informação: uma reflexão crítica. **Informação & Informação**, Londrina, v. 13, n. 2, p. 56-78, jul./dez. 2008.

SHERA, J. Epistemologia social, semântica geral e Biblioteconomia. **Ciência da Informação**, Brasília v. 6, n. 1, p. 9-12, 1977.

STRZALKOWSKI, T. (Ed.). **Natural Language Information Retrieval**, v. 7 of *Text, Speech and Language Technology*. Kluwer Academic Press: Dordrecht, Boston, London, 1999.

ZANDONADE, T. Epistemologia da Ciência da Informação. **Revista de Biblioteconomia de Brasília**, Brasília, v. 23-24, n. 3, 1999-2000.

ZOBOLI, F.; SILVA, R. I.; BORDAS, M. A. G. Cisão corpo/ mente na escola: uma análise a partir da epistemologia social. **Educação Temática Digital**, Campinas, v. 8, n. 1, p. 1-22, dez. 2006.