

## Descoberta Baseada em Literatura: Estado da Arte

Cristiane Woszezenki<sup>1</sup>  
Alexandre Leopoldo Gonçalves<sup>2</sup>

### Resumo

A Mineração de Textos (MT) pode ser entendida como o estudo e a prática de extrair informação de textos usando os princípios da linguística computacional. Uma das subáreas da mineração de texto que tem recebido grande atenção de pesquisadores de áreas como biomedicina e bioinformática é a Descoberta Baseada em Literatura (DBL). A DBL tem por objetivo encontrar relacionamentos implícitos na literatura científica de forma a gerar hipóteses potenciais para novas descobertas. Este artigo apresenta o estado da arte na DBL, abordando os modelos de descoberta utilizados, o espaço de descoberta bem como as técnicas, os métodos e as ferramentas que são empregadas na automatização do processo de descoberta.

**Palavras-chave:** Mineração de textos. Descoberta Baseada em Literatura. Literatura biomédica.

---

<sup>1</sup>Instituto Federal de Santa Catarina (IFSC) – Araranguá, cristiane.raquel@ifsc.edu.br

<sup>2</sup> Universidade Federal de Santa Catarina (UFSC) – alexandre.goncalves@ararangua.ufsc.br

## 1 INTRODUÇÃO

A Mineração de Textos (MT) pode ser entendida como o estudo e a prática de extrair informação de textos usando os princípios da linguística computacional (SULLIVAN, 2001). A MT é composta por diversas atividades, como pré-processamento (classificação) dos documentos, extração de informação, reconhecimento de entidades, extração de relacionamentos, agrupamento de textos, geração de hipóteses, entre outras. Grande parte dessas atividades é desenvolvida com o uso das técnicas de Processamento de Linguagem Natural (PLN) e análise estatística (ZHOU, 2008).

Uma das subáreas da mineração de texto que tem recebido grande atenção de pesquisadores de áreas como biomedicina e bioinformática é a Descoberta Baseada em Literatura (DBL – *Literature-Based Discovery*, também chamada de *Literature-Related Discovery* (KOSTOFF, 2008)). A DBL tem por objetivo encontrar relacionamentos implícitos na literatura científica de forma a gerar hipóteses potenciais para novas descobertas (SMALHEISER, 2011). As descobertas se dão na forma de conexão entre dois conceitos primários, como por exemplo, um medicamento para uma doença ou para um gene que é causa de uma doença (HRISTOVSKI, 2006).

A literatura biomédica é a fonte de dados mais explorada pelos trabalhos em DBL e a grande maioria das pesquisas em DBL busca encontrar novos diagnósticos, prevenções e tratamentos (COHEN, 2005). O aumento expressivo das publicações de trabalhos nesta área levou a uma fragmentação da literatura em diversas especialidades ou disciplinas com pouca comunicação entre si. Isso faz com que pesquisadores não tenham conhecimento de trabalhos relevantes de outros grupos. Assim, a DBL pode ser utilizada para fazer a conexão de diferentes disciplinas (GANIZ, 2006). Suponha que uma comunidade científica sabe que B é uma das características da doença C. Outro grupo de pesquisadores sabe que a substância A afeta B. A descoberta, nesse caso, é o levantamento da conexão implícita de A com C por meio de B (WEEBER, 2001).

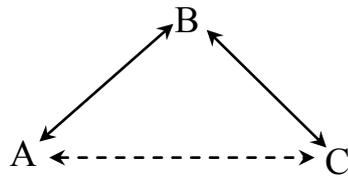
A DBL surgiu a partir das pesquisas realizadas por Don Swanson, que resultou na conexão da Síndrome de Raynaud (uma condição que resulta em restrição intermitente do fluxo sanguíneo para os dedos, disparado pelo frio ou estímulos emocionais) com o óleo de peixe (SWANSON, 1986). Inicialmente, Swanson revisou a

literatura biomédica em busca de informações sobre a Síndrome de Raynaud. Seus estudos apontaram que pacientes que possuem esta doença apresentam alterações no sangue, como “alta viscosidade” e “elevada agregação de plaquetas”. Na revisão da literatura sobre alta viscosidade do sangue, Swanson encontrou uma conexão entre esse termo e o termo “óleo de peixe”, indicando que o óleo de peixe auxilia na diminuição da viscosidade do sangue e a agregação de plaquetas. Assim, ele encontrou um relacionamento indireto entre a doença “Síndrome de Raynaud” e os fatores “alta viscosidade do sangue” e “elevada agregação de plaquetas”. Isso conduziu para a geração da hipótese que “óleo de peixe” pode ser útil para reduzir a “alta viscosidade do sangue” e a “elevada agregação de plaquetas” em seres humanos e então amenizar os sintomas da “Síndrome de Raynaud”. Mais tarde, Swanson também encontrou um relacionamento entre os termos “Magnésio” e “Enxaqueca” (SWANSON, 1988).

O primeiro trabalho realizado por Swanson, responsável por introduzir a DBL, foi realizado manualmente. Posteriormente a essa importante descoberta, muitos pesquisadores da área de ciências biomédicas e bioinformática dedicaram-se a estudar e desenvolver técnicas, métodos e ferramentas computacionais para auxiliar na geração de hipóteses potenciais para descoberta. Assim, este artigo apresenta o estado da arte da DBL, discutindo: modelos de descoberta utilizados (seção 2); espaços de descoberta, ou seja, contextos explorados pela a DBL (seção 3) e; técnicas, métodos e ferramentas que são empregadas para automatizar o processo de descoberta.

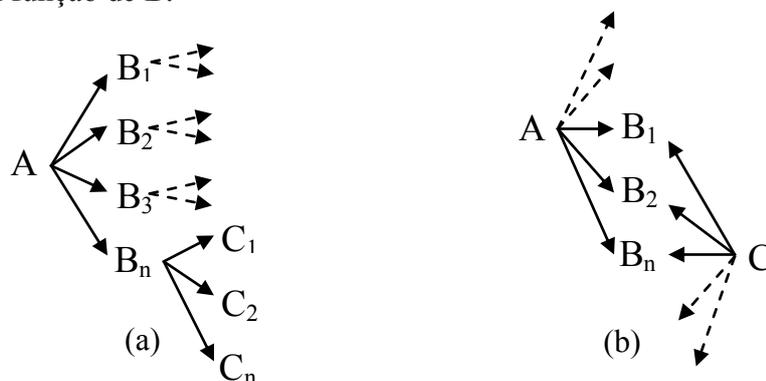
## **2 MODELOS DE DESCOBERTA**

Desde que Swanson fez a sua primeira descoberta e introduziu o conceito de DBL, o modelo por ele utilizado ainda continua sendo o mais empregado pelos trabalhos em DBL. Tal modelo é chamado de ABC (WEEBER, 2003). Por exemplo, dada a afirmação em um determinado artigo de que “A afeta B”, e dada a informação de que “B afeta C” apresentada por outro artigo, pode-se derivar a afirmação implícita de que “A afeta C” (Figura 1) (SMALHEISER, 2011). Ou seja, a premissa dessa abordagem é que existem duas disciplinas ou estruturas de conhecimento científico que não se comunicam diretamente. Contudo, parte do conhecimento de um domínio pode complementar o conhecimento do outro (WEEBER, 2003).



**Figura 1:** Modelo ABC de Swanson. Os relacionamentos AB e BC são conhecimentos e apresentados explicitamente na literatura. O relacionamento AC é uma possível descoberta (Fonte: WEEBER, 2003)

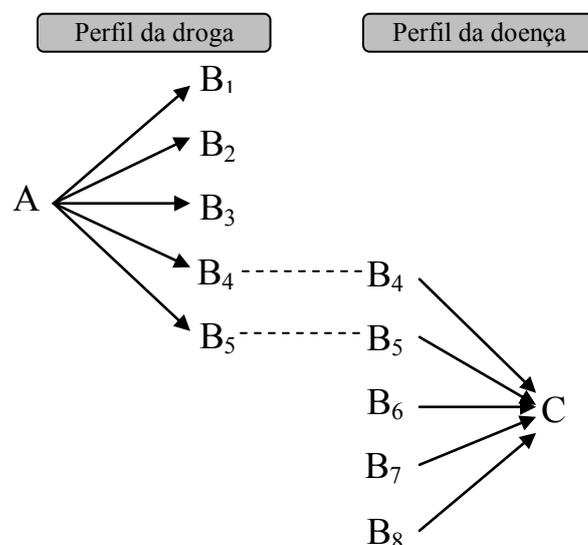
Weeberet al. (2001) definiram duas abordagens de descoberta para o modelo ABC: fechada e aberta. A Figura 2 ilustra ambas as abordagens. A descoberta aberta é caracterizada como a geração de hipóteses como resultante do processo da DBL. Seu processo inicia com um conceito (um tópico ou um termo) conhecido, denotado como conceito inicial ou A. Um exemplo é uma doença. Em seguida, este conceito é utilizado na consulta a uma base de documentos, e todos os documentos contendo esse conceito são utilizados, compondo a literatura inicial. Termos importantes ou frases podem ser extraídos e processados utilizando diferentes técnicas com o auxílio de um especialista humano. Cada termo ou frase da lista resultante é chamado de conceito intermediário ou B. Por exemplo, se A é uma doença (tal como “Doença de Raynaud”), então os conceitos B podem ser características ou sintomas da doença (como “agregação de plaquetas” e “viscosidade do sangue”). A literatura intermediária é processada e novos termos ou frases são extraídos novamente. Finalizando o processo, um conceito é selecionado por um especialista humano, denotado de conceito alvo ou C (como por exemplo, “óleo de peixe), gerando uma hipótese: C pode ser usado no tratamento da doença A em função de B.



**Figura 2:** Modelo de descoberta aberta (a) e modelo de descoberta fechada (b). As linhas contínuas indicam caminhos com potenciais de descoberta; as linhas tracejadas indicam caminhos sem sucesso. (Fonte: GANIZ, 2005)

A descoberta fechada começa com A e C conhecidos. Podem ser uma associação observada, ou uma já hipótese já gerada. A descoberta nesta situação concentra-se em encontrar novos Bs que podem explicar essa observação. É importante ressaltar que, em ambas as abordagens, tanto a aberta quanto a fechada, para que a descoberta seja efetivada, a hipótese formulada não pode ter sido explicitamente publicada.

Weeber et. al (2000) integram o modelo ABC de Swanson ao modelo de Vos (RIKKEN & VOS, 1995). O Modelo de Vos foca na interação de perfis de drogas com perfis de doenças. O perfil de uma droga consiste de todos os seus efeitos no corpo humano. Alguns desses efeitos são desejáveis, outros não. O perfil de uma doença consiste de todos os seus sinais e sintomas, ou seja, todas as características da doença. Vos define o processo de descoberta da droga em função da aproximação dessa com a doença relacionando seus perfis. Na integração desse modelo ao modelo ABC, Weeber mostra que as características dos perfis no modelo de Vos são os conceitos intermediários (ou B's) do modelo de Swanson. Ou seja, o perfil da droga A, por exemplo, pode incluir a característica terapêutica (B) de “redução da demanda de oxigênio” enquanto que “aumento da demanda do oxigênio” pode ser a característica da doença C. A Figura 3 ilustra a proposta de Weeber.

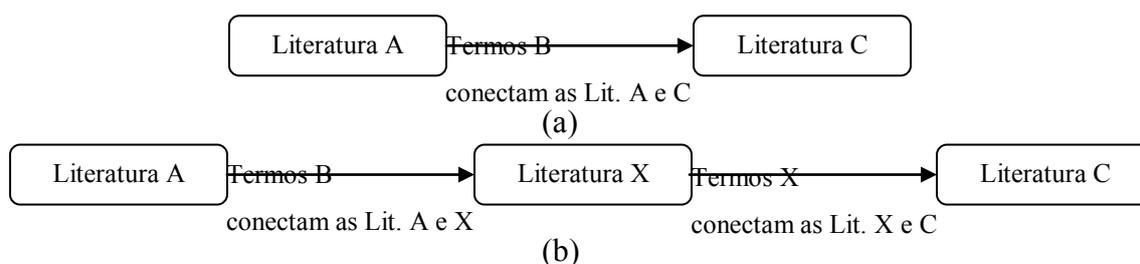


**Figura 3:** Modelo de Vo's integrado ao modelo de Swanson. (Fonte: WEEBER, 2003)

Smalheiser (2011) aponta algumas limitações do modelo ABC proposto por Swanson e ainda utilizado pela grande maioria dos trabalhos em DBL. Primeiro, o número de termos B causa uma explosão exponencial que é difícil tratar computacionalmente e requer que atalhos sejam implementados. Segundo, o grande número de termos C torna difícil analisar ou interpretar manualmente, de forma que é crucial a utilização de procedimentos que identificam os termos mais promissores. Embora diferentes sistemas tratem esse problema de formas diferentes, a grande maioria deles empregam algoritmos de similaridade, que ranqueiam os termos C de acordo com a semelhança com conceitos já conhecidos para o termo A. Dessa forma, as descobertas resultantes são *incrementais*, como por exemplo, uma determinada droga que foi desenvolvida para uma doença em específico pode ser usada no tratamento de outra doença que apresente características parecidas, como é o caso do modelo de Vos.

Ainda, Smalheiser (2011) afirma que os sistemas que utilizam o modelo de descoberta ABC geralmente são avaliados levando em conta que as afirmações AC sejam altamente prováveis. Contudo, novas descobertas geralmente apresentam-se como hipóteses muito improváveis (Simonton, 2004 apud in Smalheiser, 2011), o que, no caso de verdadeiras, levariam a descobertas radicais e inovação.

Seguindo nessa linha da descoberta radical, Kostoff (2008) propõe aumentar o número de links que compõem o caminho intermediário entre A e C. Isto porque, inserir grau(s) extra(s) de separação potencializa a descoberta, pois consegue relacionar várias literaturas de disciplinas diferentes, e seus respectivos pesquisadores passam a ter maior conhecimento de trabalhos de disciplinas afins, mas ignoradas. A Figura 4 apresenta a proposta de Kostoff comparada com o modelo de Swanson.



**Figura 4:** (a) Modelo ABC de Swanson; (b) Modelo proposto por Kostoff, que insere mais links intermediários entre dois termos. (Adaptado de Kostoff, 2008)

### 3 ESPAÇO DE DESCOBERTA

O espaço de descoberta diz respeito ao contexto explorado pela DBL. O espaço mais explorado para geração de hipóteses está no conteúdo do PubMed<sup>3</sup>, uma base de artigos científicos de acesso público criada e mantida pela Biblioteca Nacional de Medicina dos Estados Unidos (*U.S. National Library of Medicine's* – NLM). O PubMed compreende mais de 21 milhões de citações da literatura biomédica e fornece acesso ao MEDLINE, uma base (também mantida pela NLM) de citações e resumos nas áreas da medicina, enfermagem, odontologia, medicina veterinária, biologia, bioquímica, evolução molecular, entre outros. O MEDLINE<sup>4</sup> compreende mais de 18 milhões de publicações de aproximadamente 5.400 periódicos biomédicos do mundo inteiro desde 1948. Um registro no MEDLINE inclui o título, o autor e a informação de publicação de um artigo. Ele pode também conter o resumo do artigo.

Dentro das Ciências Biomédicas, alguns espaços de descoberta devem ser formulados de maneira mais específica, como é o caso da Medicina Tradicional Chinesa (MTC), cuja filosofia reflete as crenças da sociedade chinesa, definindo assim metodologias e abordagens diferentes para diagnóstico e tratamento (ZHOU, 2008). Nesse contexto, Zhou (2008) fez um estudo sobre a mineração de textos para a descoberta de conhecimento na MTC. Para a criação do seu espaço de descoberta, foram utilizados documentos da base de dados da CNKI (*China National Knowledge Infrastructure*, <http://www.global.cnki.net/>), uma das maiores bases de publicações científicas chinesas. Publicações internacionais foram selecionadas dos periódicos Elsevier, Springer, ACM e IEEE. Consultas ao PubMed também foram realizadas de forma a obter publicações sobre a MTC.

Na tentativa de verificar se a DBL pode ser bem sucedida em áreas diferentes da biomédica, Kostoff (2008) apresentou um estudo com o objetivo de identificar potenciais melhorias para o processo de purificação da água, especialmente no que diz respeito à redução de custos. O espaço de descoberta é formado por dois conjuntos de literatura: a literatura núcleo, que reúne documentos sobre a purificação da água; e a

---

<sup>3</sup>Pub PubMed, <http://www.ncbi.nlm.nih.gov/pubmed/>

<sup>4</sup>MEDLINE, <http://www.nlm.nih.gov/pubs/factsheets/medline.html>

literatura expandida, que reúne documentos cujos termos, conceitos e frases estão direta ou indiretamente relacionados aos conceitos na literatura núcleo.

#### **4 FERRAMENTAS, MÉTODOS E TÉCNICAS PARA AUTOMATIZAÇÃO DA DBL**

Os primeiros trabalhos desenvolvidos por Swanson (1986; 1988) foram realizados manualmente, com pesquisa extensiva na literatura, lendo muitos títulos e resumos de publicações científicas. A partir disso, Swanson, juntamente com Smalheiser, passou a utilizar técnicas computacionais de análise de texto para automatizar o processo de descoberta, baseando-se em suas estratégias anteriores. Essas técnicas levaram à criação da ferramenta de suporte à descoberta, Arrowsmith, disponível na web<sup>5</sup> (SWANSON, 1997; SMALHEISER, 2006). O processo de descoberta consiste na seleção de dois conjuntos de literatura (literatura “A” e literatura “C”); e na intersecção dos dois conjuntos e criação de uma lista de termos de interesse potencial (termos B), apresentados por ordem de relevância; edição da lista de termos B pela interação do usuário.

A ferramenta Arrowsmith utiliza o MEDLINE como espaço de descoberta. Os registros do MEDLINE são indexados de acordo com o vocabulário MeSH<sup>6</sup> (*Medical SubjectHeadings*), que pode ser entendido também como um tesouro. Esse vocabulário fornece uma forma consistente de recuperar informação, uma vez que diferentes formas terminológicas podem ser usadas para descrever o mesmo conceito. Vários outros trabalhos utilizam o vocabulário MeSH para a recuperação de conceitos (HRISTOVSKI, 2005)

Outra ferramenta de suporte a descoberta baseada em literatura biomédica é o BITOLA<sup>7</sup> (HRISTOVSKI et. al, 2005). De forma a tornar o sistema mais adequado a descoberta de genes candidatos a doenças, os autores integram conhecimento genético sobre a localização cromossômica da doença inicial bem como a localização cromossômica dos genes candidatos. Assim, embora a fonte principal de informações seja o MEDLINE, essa ferramenta utiliza as bases LocusLink (da NML) e HUGO

---

<sup>5</sup>Arrowsmith, [http://arrowsmith.psych.uic.edu/arrowsmith\\_uic/index.html](http://arrowsmith.psych.uic.edu/arrowsmith_uic/index.html)

<sup>6</sup>MeSH, <http://www.ncbi.nlm.nih.gov/mesh>

<sup>7</sup>Bitola, <http://ibmi.mf.uni-lj.si/bitola/>

(*HumanGenomeOrganization*) para o mapeamento adequado de nomes de genes e localizações cromossômicas.

No processo de geração de hipóteses da DBL, devem ser estabelecidos relacionamentos entre os conceitos biomédicos, como genes, proteínas, doenças, drogas, entre outros. Os métodos mais comuns utilizados na DBL para estabelecer esses relacionamentos são baseados na coocorrência (SWANSON & SMALHEISER, 1997; KOSTOFF, 2008; HRISTOVSKI, 2005; WEEBER, 2000; SRINIVASAN, 2004; GORDON & LINDSAY (1996); FULLER, 2004; VAN DER EIJK et al (2004); GORDON & DUMAIS, 1998; BAKER & HEMMINGER, 2010; TSURUOKA et al, 2011). Esses métodos partem do pressuposto de que é possível identificar estatisticamente um possível relacionamento entre palavras, analisando suas frequências individuais e conjuntas. Esse conceito pode ser generalizado para ser usado com o objetivo de se determinar a força do relacionamento entre dois elementos textuais que aparecem conjuntamente em algum documento textual (GONÇALVES, 2006).

De acordo com Hristovski (2006), a coocorrência não fornece nenhuma informação semântica sobre a natureza das relações entre os conceitos. O autor afirma que o uso da coocorrência tem diversas desvantagens, uma vez que nem todas as coocorrências implicam em relações interessantes: (a) nem todas as coocorrências representam uma relação; (b) os usuários devem ler um grande número de citações de forma a verificar as relações candidatas; (c) os sistemas tendem a produzir um grande número falsas relações; e (d) não existe explicação explícita da relação descoberta. Assim, ele propõe o uso de técnicas de PLN (Processamento de Linguagem Natural) (HRISTOVSKI, 2006) para melhorar a relação sinal-ruído, capturando relações semânticas entre os conceitos e explicitando sentenças na forma “A afeta/vincula-se a/regula/interage com B” e “B afeta/vincula-se a/regula/interage com C”.

Segundo Zweigenbaum (2009), a PLN fornece a base para uma mineração de texto mais sofisticada: derivação de informações do texto por meio de afirmações e associações entre entidades, e inferência, que serve como fundamentação e sumarização, respostas a questões, predição e descoberta, bem como mesclar automaticamente o conhecimento derivado de informações textuais com informações extraídas de bases de conhecimento. Diversos trabalhos utilizam técnicas de PLN para o processo de descoberta do conhecimento (ÖZGUR et. al, 2010) .

## 5 CONSIDERAÇÕES FINAIS

A Descoberta Baseada em Literatura tem se tornado cada vez mais o foco da pesquisa da descoberta de conhecimento em bases textuais. A DBL é uma atividade centrada em humanos assistidos por um programa de computador. Embora vários progressos tenham sido conquistados em direção a um processo de DBL automatizado, ainda existe muito trabalho a ser realizado, pois os sistemas atuais requerem processo manual com interações humanas em diversos pontos.

Nesta revisão de trabalhos sobre DBL, percebe-se que existe uma concentração muito grande de esforços na área biomédica. Contudo, a DBL pode ser aplicada com grande potencial em qualquer área do conhecimento.

### **Literature-Based Discovery: State of the Art**

#### **Abstract**

Text Mining (TM) can be understood as the study and practice of extracting information from text using the principles of computational linguistics. One of the sub-areas of text mining that has received great attention from researchers in areas such as biomedicine and bioinformatics is the Literature-Based Discovery (LBD). LBD aims at finding relationships implicit in the scientific literature to generate hypotheses for potential new discoveries. This article discusses the models used in the discovery LBD, techniques, methods and tools that are used in automating the process of discovery.

**Keywords:** Text Mining. Literature Based Discoveries. Biomedical Literature.

#### **REFERÊNCIAS**

BAKER, N. C.; HEMMINGER B. M. Mining connections between chemicals, proteins, and diseases extracted from Medline annotations. **Journal of Biomedical Informatics**.43(4), p. 510-519, 2010.

COHEN, A. M., HERSH, W. R. (A Survey of Current Work in Biomedical Text Mining. **Briefings in Bioinformatics**, Vol. 6, Issue 1, p. 57-71, 2005.

FULLER, S. S.; REVERE, D., BUGNI, P. F.; MARTIN, G. M. A knowledge base system to enhance scientific discovery: Telemakus. **Biomed. Digit Libr.**, Vol. 1(1), p. 2, 2005.

GANIZ, M. C.; POTTENGER, W. M.; JANNECK, C. D. Recent Advances in Literature Based Discovery. **Journal of the American Society for Information Science and Technology**, JASIST, 2006.

GONÇALVES, A. L. **Um modelo de descoberta de conhecimento baseado na correlação de elementos textuais e expansão vetorial aplicado à engenharia e gestão do conhecimento**. 196 (Doutorado). Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis (SC), 2006.

GORDON, M. D.; DUMAIS, S. Using Latent Semantic Indexing for literature based discovery. **Journal of the American Society for Information Science and Technology**, 49(8), p. 674-685, 1998.

GORDON M.D.; LINDSAY R.K. Toward discovery support systems: A replication, re-examination, and extension of Swanson's work on literature-based discovery of a connection between Raynaud's and fish oil. **J Am SocInfSci** 1996; 47(2), p.116-128, 1996.

HRISTOVSKI, D.; FRIEDMAN, C.; RINDFLESCH T. C.; PETERLIN, B. Exploiting Semantic Relations for Literature-Based Discovery. **AMIA Annual Symposium Proceedings**, p. 349–353, 2006.

HRISTOVSKI, D.; PETERLIN, B.; MITCHELL, J. A.; HUMPHREY, S. M. Using literature-based discovery to identify disease candidate genes”. **Int J Med Inform** 74(2–4), p. 289–298, 2005.

KOSTOFF, R. N.; BRIGGS, M. B.; SOLKA, J. L.; RUSHENBERG, R. L. Literature-related discovery (LRD): Methodology. **Technological Forecasting and Social Change**, Vol. 75, Issue 2, p. 186-202, 2008.

RIKKEN, F.; VOS, R. How adverse drug reactions can play a role in innovative drug research. **Pharmacy World and Science**, 17(6), p. 195–200, 1995.

SMALHEISER N. R.; TORVIK V. I.; BISCHOFF-GRETHER A.; BURHANS L. B.; GABRIEL M.; HOMAYOUNI R.; et al. Collaborative development of the Arrowsmith two node search interface designed for laboratory investigators. **J Biomed DiscovCollab** 1(8), 2006.

SMALHEISER, N. R. Literature-Based Discovery: Beyond the ABCs. **Journal of the American Society for Information Science and Technology**.doi: 10.1002/asi.21599, 2011.

ÖZGÜR, A.; XIANG, Z.; RADEV, D. R.; HE, Y. Literature-Based isDcovery of IFN- $\gamma$  and Vaccine-Mediated Gene Interaction Networks”. **Journal of Biomedicine and Biotechnology**. Volume 2010, Article ID 426479, 13 pages, doi:10.1155/2010/426479, 2010.

SRINIVASAN, P.; LIBBUS, B. Mining MEDLINE for implicit links between dietary substances and diseases. **Bioinformatics**, Vol. 20, Suppl 1, p. I290–I296, 2004.

SULLIVAN, D. **Document Warehousing and Text Mining**. New York: John Wiley & Sons, Inc, 2001.

SWANSON, D. R. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. **Perspectives in Biology and Medicine**, v. 30(1), p. 7-18, 1986.

\_\_\_\_\_. Migraine and Magnesium: eleven neglected connections. **Perspectives in Biology and Medicine**.v. 31(4), p. 526-557, 1998.

SWANSON D. R.; SMALHEISER, N. R. An interactive system for finding complementary literatures: a stimulus to scientific discovery. **Artificial Intelligence**.91, p. 183-203, 1997.

TSURUOKA, Y.; MIWA, M.; HAMAMOTO, K.; TSUJII, J.; ANANIADOU, S. Discovering and visualizing indirect associations between biomedical concepts. **Bioinformatics**.Vol. 27, Issue13, p. i111-i119, 2011.

VAN DER EIJK, C.; VAN MULLIGEN, E.; KORS, J.A.; MONS, B.; VAN DEN BERG, J. Constructing an associative concept space for literature-based discovery. **Journal of the American Society for Information Science and Technology**, 55(5), p. 436-444, 2004.

WEEBER M.; KLEIN H.; ARONSON A.R.; MORK J.G.; JONG-VAN DEN BERG L.; VOS R. Text-based discovery in biomedicine: the architecture of the DAD-system. **Proc AMIA Symp**.2000;(20 Suppl):903-7, 2000.

WEEBER, M. et al. Using concepts in literature-based discovery: simulating Swanson's Raynaud-fish oil and migraine-magnesium discoveries. **Journal of the American Society for Information Science and Technology**, v. 52, n. 7, pp. 548-557. ISSN 1532-2882, 2001.

WEEBER, M. Advances in Literature-Based Discovery. In: **Journal of the American Society for Information Science and Technology**. 2003.

ZHOU, X.; PENG, Y.; LIU, B. Text mining for traditional Chinese medical knowledge discovery: A survey. **Journal of Biomedical Informatics**, Vol. 43, Issue 4, p. 650-660, 2008.

ZWEIGENBAUM, P.; DEMNER-FUSHMAN D. Advanced Literature-Mining Tools. **Bioinformatics: Tools and Applications**, 2009.