



Ministério da
Ciência e Tecnologia



INPE-16620-RPQ/8384

UM SISTEMA DE RECOMENDAÇÃO DE PUBLICAÇÕES CIENTÍFICAS BASEADO EM AVALIAÇÃO DE CONTEÚDO

Alessandro Oliveira Arantes

Relatório final da disciplina Princípios e Aplicações de Mineração de Dados
(CAP-359) do Programa de Pós-Graduação em Computação Aplicada, ministrada
pelo professor Rafael Santos.

Registro do documento original:

<<http://urlib.net/sid.inpe.br/mtc-m18@80/2009/10.01.19.32>>

INPE
São José dos Campos
2009

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3945-6911/6923

Fax: (012) 3945-6919

E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO:

Presidente:

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Membros:

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Haroldo Fraga de Campos Velho - Centro de Tecnologias Especiais (CTE)

Dr^a Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Dr. Ralf Gielow - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr. Wilson Yamaguti - Coordenação Engenharia e Tecnologia Espacial (ETE)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Jefferson Andrade Ancelmo - Serviço de Informação e Documentação (SID)

Simone A. Del-Ducca Barbedo - Serviço de Informação e Documentação (SID)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Marilúcia Santos Melo Cid - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Viveca Sant´Ana Lemos - Serviço de Informação e Documentação (SID)

SUMÁRIO

	<u>Pág.</u>
LISTA DE FIGURAS	
LISTA DE TABELAS	
1 INTRODUÇÃO	9
2 MATERIAIS E MÉTODOS.....	10
2.1 Sistema de busca	12
2.2 Sistema de recomendação	15
2.3 Análise de heurística empregada	15
3 CONCLUSÕES	23
REFERÊNCIAS BIBLIOGRÁFICAS.....	25

LISTA DE FIGURAS

	<u>Pág.</u>
2.1 - Página de entrada da aplicação com campo de busca.....	13
2.2 - Lista de ocorrências encontradas na busca	14
2.3 - Página de detalhes da publicação selecionada	15
2.4 - Tabela de pontos entre os relacionamentos das publicações.....	17
2.5 - Recurso que possibilita a visualização dos pontos de cada sugestão	18
2.6 - Todas as ligações para a publicação CTA/IEAv-EFO/AP-004/2005	19
2.7 - Somente ligações com 3 ou mais pontos.....	20
2.8 - Ligações com 3 ou mais pontos utilizando dois níveis de busca	21
2.9 - Grafo Ligações com 5 ou mais pontos utilizando dois níveis de busca	22

LISTA DE TABELAS

	<u>Pág.</u>
2.1 - Tabela de regras e pontos aplicados nas recomendações	16

1 INTRODUÇÃO

O CTA (Comando-Geral de Tecnologia Aeroespacial) é um instituto de pesquisa da aeronáutica que tem uma relevante parte de seu patrimônio intelectual presente nas inúmeras publicações técnico científicas. Tais publicações estão diretamente ligadas aos projetos e atividades do instituto, e além de estimularem o compartilhamento de conhecimento em pesquisa, também são indicativos práticos de sua produtividade bem como um meio de divulgação dos trabalhos desenvolvidos para a comunidade científica e até mesmo para a sociedade como um todo.

No ano de 1997 a divisão de desenvolvimento de sistemas do IEAv (Instituto de Estudos Avançados) desenvolveu um sistema de cadastro e *workflow* de publicações científicas baseado na arquitetura cliente-servidor do IBM Lotus Notes. Este sistema foi desenvolvido baseado numa norma de padronização das informações das publicações e foi utilizado como forma de armazenamento digital das mesmas por praticamente todos os institutos do CTA até 2005. De 2005 em diante somente o IEAv continuou utilizando esse sistema, devido à decisões internas do CTA, e este se encontra operante até hoje contando com aproximadamente 3.000 publicações cadastradas. Os tipos de publicações que foram inseridas no sistema são:

- a) Artigos de Evento Científico
- b) Artigos de Periódico
- c) Capítulos de Livro
- d) Livro
- e) Manual Técnico
- f) Nota Técnica
- g) Relatório de Pesquisa
- h) Trabalho de Graduação
- i) Dissertações de Mestrado

j) Teses de Doutorado

Este sistema foi desenvolvido à mais de uma década atrás, portanto obviamente não contava com muitos recursos para agregar funcionalidades à aplicação devido às limitações da plataforma. Com isso, atualmente uma mudança de plataforma está sendo estudada em virtude do avanço nos últimos anos principalmente de plataformas gratuitas, já que a utilização de software gratuito é uma recomendação da própria aeronáutica. Conseqüentemente, um novo sistema de publicações deverá ser desenvolvido à partir de 2010 e, além de ser disponibilizado para acesso via internet, ao invés de se limitar aos usuários do IBM Lotus Notes, dois bons exemplos de funcionalidades que certamente agregariam valor ao sistema são um sistema de busca de publicações para a internet através de palavra(s) chave(s) e a capacidade de recomendar ao usuário algumas publicações de conteúdo semelhante à que ele visualiza, também via internet. Como motivação para o trabalho do curso de “Princípios e Aplicações de Mineração de Dados”, ambos recursos deverão ser implementados e analisados no contexto do foco da disciplina.

2 MATERIAIS E MÉTODOS

Atualmente as informações das publicações que estão armazenadas em Banco de Dados do Lotus Notes não podem ser acessadas através de uma linguagem de programação que possa ser utilizada para montagem de páginas na internet, portanto as informações das publicações terão que ser exportadas, inseridas em Banco de Dados MySQL, pré-processadas, e acessadas através de uma interface implementada em PHP (*Hypertext Pre-processor*).

Resumindo, podemos dizer que o aplicação será dividida em três partes:

- a) Implementação do sistema de busca: Provavelmente a tarefa mais simples à ser executada, consiste basicamente em disponibilizar um campo onde os usuários poderão digitar a(s) palavra(s) chave(s) e a

mesma será buscada em campos pré determinados do banco de dados retornando resultados em ordem das publicações mais recentes para as mais antigas;

- b) Implementação do sistema de recomendação: Para que um sistema possa recomendar publicações de conteúdo semelhante através da mineração do banco de dados, uma heurística deverá ser pensada e implementada de forma que possa ser utilizada pelos usuários em tempo hábil fornecendo resultados satisfatórios.
- c) Conclusão com análise dos resultados das recomendações e da heurística utilizada: A heurística utilizada trouxe resultados satisfatórios? Outros métodos poderiam ser utilizados de forma à otimizar os resultados? Será feita uma análise sobre a execução desse trabalho.

Um sistema de busca simples será o mecanismo de início deste trabalho, pois através dele o usuário fará o acesso à uma determinada publicação e, baseado numa avaliação do conteúdo dessa publicação em relação às demais, o sistema fará a recomendação de até 5 (cinco) outras novas publicações que poderão interessar ao usuário devido à semelhança de conteúdo. Essa semelhança entre as publicações deverão ser determinadas através dos campos de informações disponíveis que são:

Ano: ano de publicação.

Tipo: tipo da publicação (Artigo de Periódico, Livro, Tese de Doutorado, etc.)

Número: identificação única da publicação na biblioteca digital.

Título: título da publicação.

Resumo: resumo do conteúdo da publicação.

Periódico/Evento: nome do periódico (se artigo de periódico) ou evento (se artigo de evento) da publicação.

Título do Livro: título do livro, caso a publicação seja um capítulo de livro.

Autor: autor do livro, caso a publicação seja um capítulo de livro.

Instituição: nome da instituição à qual a publicação se refere (se trabalho de graduação, dissertação ou tese).

Faculdade: nome da faculdade (se trabalho de graduação).

Área: área de trabalho da publicação (se trabalho de graduação, dissertação ou tese).

Local: local da publicação (no Brasil ou exterior).

Palavras-Chaves: são compostas de palavras-chaves sugeridas pelos autores, e termos chaves que são escolhidos numa tabela de termos padrão para publicações da aeronáutica.

Nomes dos autores: nomes de até 10 autores.

Locais de trabalho dos autores: locais de trabalho de todos os autores.

Um pré-processamento do Banco de Dados deverá ser feito para tratar problemas de erros no preenchimento, repetição de palavras-chaves e alguns campos ausentes. Esta etapa é muito importante, pois a qualidade dos dados coletados é um requisito que está diretamente ligado à capacidade de se minerar conhecimento confiável dos mesmos.

2.1 Sistema de busca

Para este fim, o preenchimento de um simples formulário acionará um script de consulta SQL (*Structured Query Language*) implementado em PHP. No instante em que o usuário digitar uma palavra-chave no campo e submeter o formulário, a palavra-chave será procurada em todos os registros do Banco de Dados e suas ocorrências serão listas abaixo em forma de link para a visualização em detalhes de cada documento. O protótipo da tela de busca pode ser observado na Figura 2.1, a lista com as ocorrências pode ser vista na Figura 2.2, e na Figura 2.3 é mostrada a página que exhibe a publicação em detalhes.

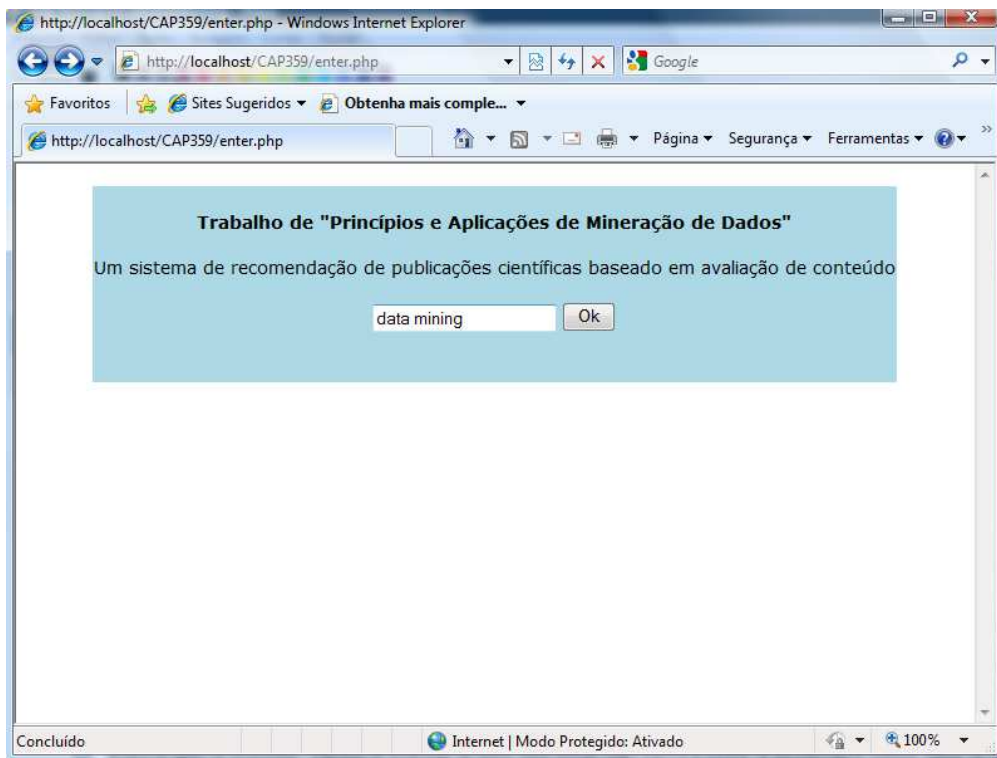


Figura 2.1 – Página de entrada da aplicação com campo de busca.

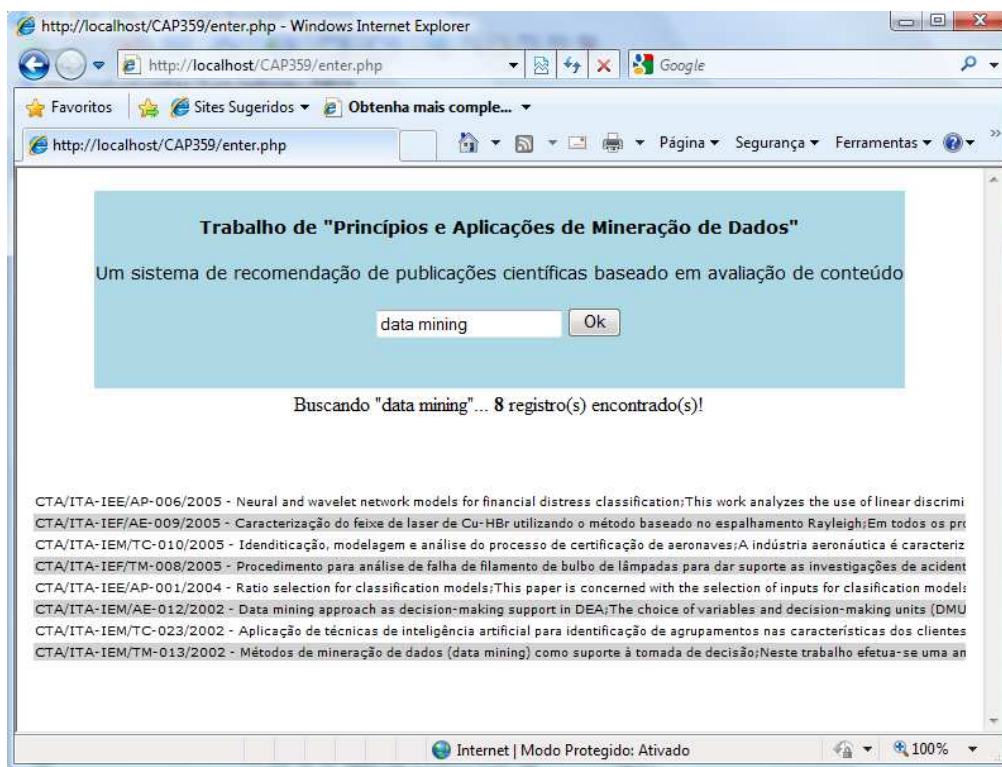


Figura 2.2 – Lista de ocorrências encontradas na busca.

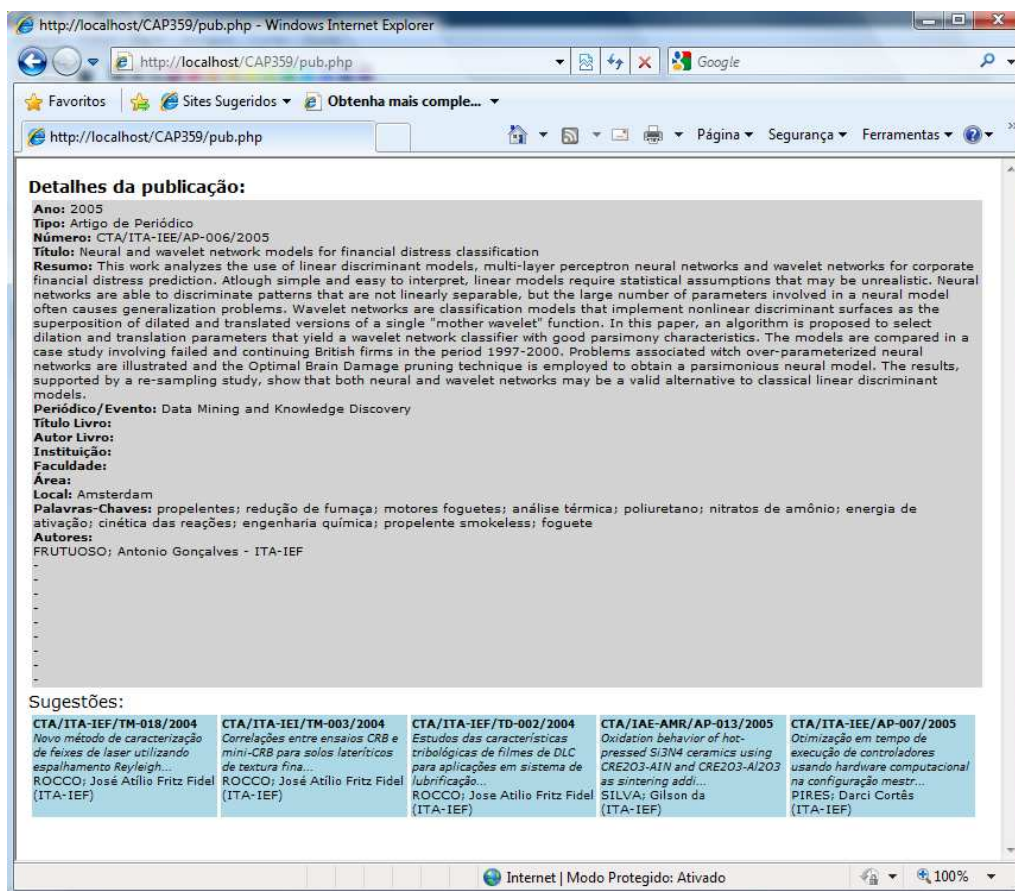


Figura 2.3 – Página de detalhes da publicação selecionada.

2.2 Sistema de recomendação

A utilização de sistemas de recomendação tem sido uma abordagem eficiente para reduzir o esforço do usuário com a finalidade de se encontrar informações de interesse (WANG et al., 2008).

Recentemente, há muitas propostas de sistemas de recomendação, sejam baseados em conteúdo, através de modelos de clusterização, regras de associação ou modelos baseados em grafos (WANG et al., 2008). Neste trabalho, conforme já mencionado, o sistema de recomendação deverá selecionar um conjunto pequeno de publicações que mais se assemelham à

publicação visualizada no momento e sugeri-las ao usuário como possíveis publicações de interesse. Vários métodos foram analisados em Kim et al. (2005); Pazzani; Billsus (2007); Jeong; Cho (2004); Wang et al. (2008); Du et al. (2007); Smyth (2007); Burke (2007), e o método que melhor se adaptou ao contexto abordado neste trabalho foi através de uma heurística de avaliação do conteúdo com o intuito de mensurar o quanto uma publicação se assemelha às outras do sistema. Essa medida que irá qualificar a proximidade entre os assuntos das publicações será feita unicamente baseada nos seus conteúdos, e para essa tarefa o campo de palavras-chaves representa uma facilidade para o desenvolvimento já que ele foi preenchido pelos próprios autores e, teoricamente, são as pessoas que mais conhecem o conteúdo de suas publicações.

Assim sendo, haverá um número inteiro que representará o resultado da comparação entre o conteúdo de duas publicações. Tais números serão nomeados de “pontos”, e a quantidade de pontos entre duas publicações indicará o quanto seus conteúdos se assemelham. Para este fim, todos os pontos serão zerados e incrementados à partir de um conjunto de regras (mostrado na Tabela 2.1) que serão aplicadas para cada ligação entre duas publicações diferentes do sistema:

Tabela 2.1 – Tabela de regras e pontos aplicados nas recomendações.

Regras	Pontos
Se forem do mesmo autor	+ 3
Se o autor da 1ª for um dos autores da 2ª	+ 2
Se os primeiros autores forem do mesmo local	+ 2
Se o autor da 1ª for de um dos locais da 2ª	+ 1
Palavras-chaves	+ 1 (para cada palavra em comum)
Publicação mais recente	+ 1
Publicação mais antiga	- 1

O armazenamento dos pontos no banco de dados será em uma tabela que conterà os pontos somados para todas as ligações entre publicações. Visando uma melhor performance, serão armazenadas somente as tuplas cuja pontuação seja maior do que 0 (zero). Em outras palavras, para n publicações teremos no máximo $(n*n)-n$ tuplas com os números de identificação das publicações de origem e destino, e os pontos correspondentes à essa ligação, conforme mostra a Figura 2.4:

de	pontos ▼	para
CTA/IEAv-EFO/RP-002/2005	11	CTA/ITA-IEA/TC-021/2005
CTA/ITA-IAB/NT-001/2005	9	CTA/IAE-AMR/RP-002/2005
CTA/IAE-AMR/RP-001/2005	9	CTA/IAE-AMR/RP-002/2005
CTA/IAE-AMR/RP-002/2005	9	CTA/IAE-AMR/RP-001/2005
CTA/ITA-IEA/TC-021/2005	9	CTA/IEAv-EFO/RP-002/2005
CTA/ITA-IAB/NT-001/2005	8	CTA/IAE-AMR/RP-001/2005
CTA/IAE-AMR/RP-001/2005	8	CTA/ITA-IAB/NT-001/2005
CTA/ITA-IEE/AE-023/2005	7	CTA/ITA-IEE/AE-025/2005
CTA/ITA-IEE/AE-025/2005	7	CTA/ITA-IEE/AE-023/2005
CTA/IAE-AMR/RP-002/2005	7	CTA/ITA-IAB/NT-001/2005
CTA/IEAv-EFO/RP-002/2005	6	CTA/ITA-IEA/TC-020/2005
CTA/ITA-IEA/TC-020/2005	5	CTA/ITA-IEA/TC-021/2005
CTA/ITA-IEA/TC-021/2005	5	CTA/ITA-IEA/TC-020/2005
CTA/ITA-IEA/TC-020/2005	4	CTA/IEAv-EFO/RP-002/2005
CTA/ITA-IAB/NT-001/2005	3	CTA/IEAv-EFO/RP-002/2005
CTA/ITA-IAB/NT-001/2005	3	CTA/ITA-IEA/TC-021/2005
CTA/IAE-AMR/RP-002/2005	3	CTA/IEAv-EFO/RP-002/2005
CTA/IAE-AMR/RP-002/2005	3	CTA/ITA-IEA/TC-021/2005
CTA/IEAv-EFO/RP-002/2005	3	CTA/ITA-IAB/NT-001/2005
CTA/IEAv-EFO/RP-002/2005	3	CTA/IAE-AMR/RP-002/2005
CTA/ITA-IEA/TC-021/2005	3	CTA/ITA-IAB/NT-001/2005
CTA/ITA-IEA/TC-021/2005	3	CTA/IAE-AMR/RP-002/2005
CTA/ITA-IEE/AE-023/2005	2	CTA/IAE-AMR/RP-001/2005
CTA/ITA-IEE/AE-025/2005	2	CTA/IAE-AMR/RP-001/2005
CTA/ITA-IAB/NT-001/2005	2	CTA/ITA-IEA/TC-020/2005

Figura 2.4 – Tabela de pontos entre os relacionamentos das publicações.

Inserir tantas tuplas no banco de dados consome tempo, para 3000 publicações poderíamos ter até 8.997.000 tuplas. Isso torna o sistema de recomendação inviável para executar no mesmo tempo em que o usuário visualiza uma publicação, por isso essas tuplas terão que ser geradas de forma independente da interface do sistema executando scripts de tempos em tempos agendados pelo sistema operacional do servidor. Dessa forma, quando um usuário tem a visão de uma determinada publicação, as sugestões relacionadas não precisam ser calculadas e inseridas no banco de dados, acontecerá apenas uma busca pelas publicações que somam mais pontos à partir da que este sendo visualizada no momento, e isso torna viável a utilização do sistema mesmo que para uma massa grande de dados.

Para facilitar a compreensão do motivo pelos quais as publicações foram sugeridas, além de mostrar ao usuário as sugestões em si, a aplicação também mostra a razão delas serem as mais relacionadas à publicação em visualização expondo as regras que foram utilizadas para somar ou subtrair pontos. Dentro dessa descrição das regras utilizadas, é possível ver se os pontos foram somados devido à semelhança de nomes de autores, locais de trabalho, comparação de palavras-chaves, e até pontos somados para publicações mais recentes. A Figura 2.5 mostra em detalhes esse recurso:

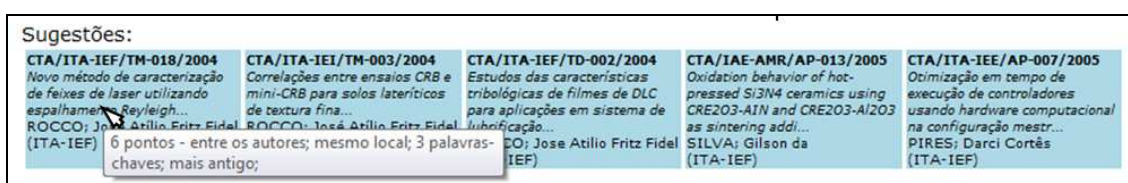


Figura 2.5 – Recurso que possibilita a visualização dos pontos de cada sugestão.

2.3 Análise da heurística empregada

As regras utilizadas e suas pontuações foram determinadas de forma à sugerir prioritariamente publicações com assunto semelhante através da comparação

entre palavras-chaves e, de forma secundária, as publicações do mesmo autor. Como último recurso, a sugestão é de publicações cujo local de trabalho dos autores são os mesmos. Também, são considerados mais pontos para publicações mais recentes favorecendo a indicação de trabalhos mais atuais em detrimento dos mais antigos.

É possível modificar sensivelmente o comportamento do sistema mudando as pontuações dadas para cada regra. Seria possível, por exemplo, favorecer as sugestões de trabalhos do mesmo autor dando mais pontos para essa regra. Alguns testes foram feitos mudando os valores agregados à cada regra e observou-se que as palavras-chaves realmente representam de forma concreta os temas abordados nos trabalhos, e por isso, foi decidido utilizar esse campo como principal fonte de pontuação.

Assumindo como estudo de caso uma busca pelo termo “*semiconductor*” no sistema, das 15 ocorrências resultantes será visualizado a primeira delas que é um Artigo de Periódico do IEAv cujo número é “CTA/IEAv-EFO/AP-004/2005” e está com o título de “*Quantum ballistic conductance of quasi-two-dimensional and three-dimensional semiconductor nanowires*”. Utilizando o JUNG (*Java Universal Network/Graph Framework*) é possível demonstrar graficamente as pontuações dos relacionamentos entre as publicações.

Na Figura 2.6 é possível ver um grafo onde estão todos os documentos ligados à publicação em questão, são 216 publicações relacionadas (ou 216 vértices) e 217 relacionamentos (arestas). O vértice vermelho representa o artigo visualizado, os vértices amarelos são as outras publicações relacionadas. A largura da ligação representa o número de pontos entre as duas publicações.

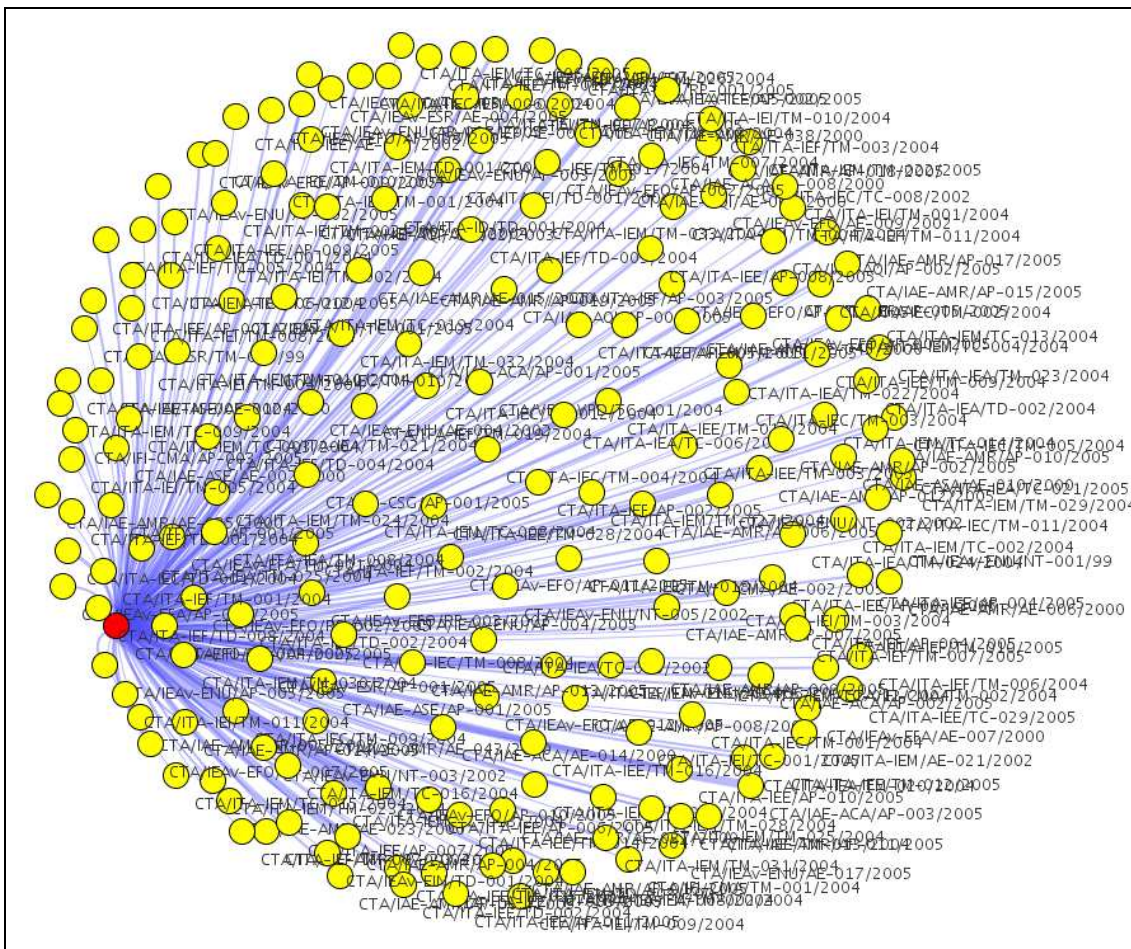


Figura 2.6 – Todas as ligações para a publicação CTA/IEAv-EFO/AP-004/2005.

Devido ao grande número de relacionamentos a visualização se torna confusa, portanto será mostrada também a Figura 2.7 que mostra apenas os relacionamentos cujos pontos sejam maiores ou iguais à 3 (três), assim é possível reduzir para um grafo com 31 vértices e 32 arestas

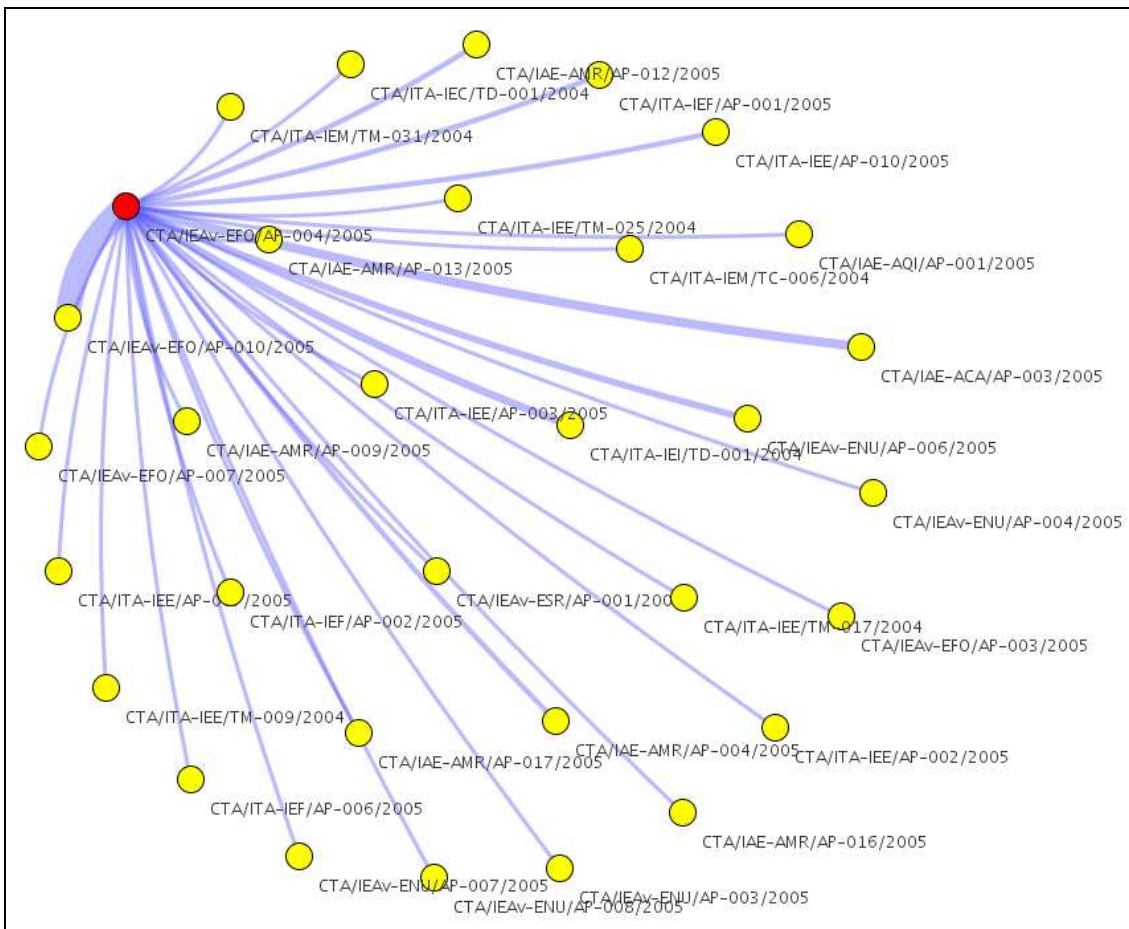


Figura 2.7 – Somente ligações com 3 ou mais pontos.

Agora sim é possível observar, por exemplo, que a publicação em questão é fortemente ligada à de número CTA/IEAv-EFO/AP-010/2005 já que ambas tem 17 palavras-chaves em comum e são oriundas do mesmo instituto.

A Figura 2.8 faz uma seleção como a Figura 2.7, só publicações com pontuação maior ou igual à 3 (três), mas apresenta dois níveis de relacionamentos. O ponto vermelho é a publicação visualizada, os pontos amarelos são as 31 publicações relacionadas diretamente à primeira, e os outros 143 pontos verdes são as publicações relacionadas às 31 resultando num grafo com 175 vértices e 1611 arestas no total.

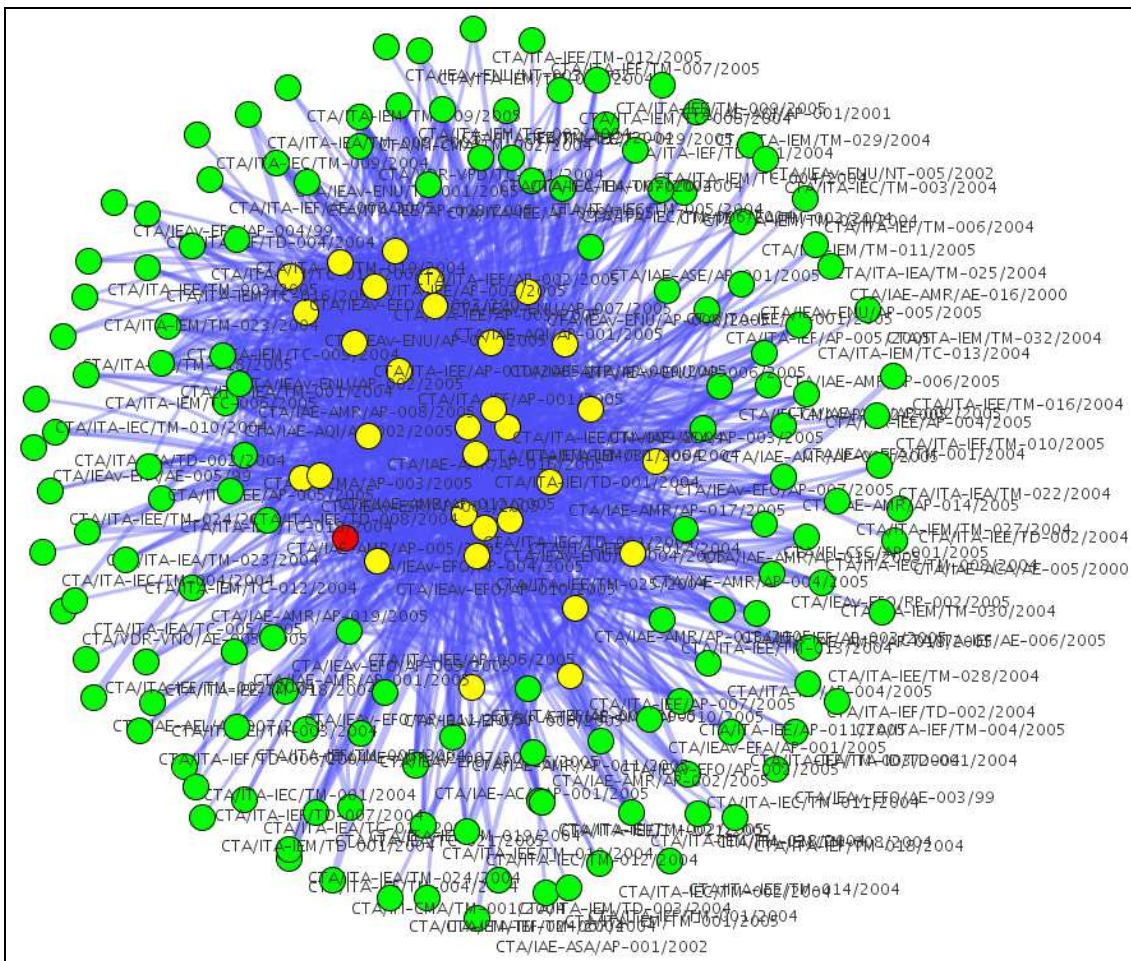


Figura 2.8 – Ligações com 3 ou mais pontos utilizando dois níveis de busca.

A Figura 2.8 não está clara devido ao grande número de vértices e arestas, a aplicação de uma seleção mais restritiva certamente reduziria consideravelmente o número de vértices e arestas. Tendo em vista essa possibilidade, a Figura 2.9 mostra em dois níveis os relacionamentos entre as publicações utilizando somente pontuações maiores ou iguais à 5 (cinco). Desta forma, além do documento de origem (vértice mais escuro), foram encontrados 4 documentos relacionados diretamente, e 18 documentos relacionados aos 4 resultando num grafo de 23 vértices e 41 arestas.

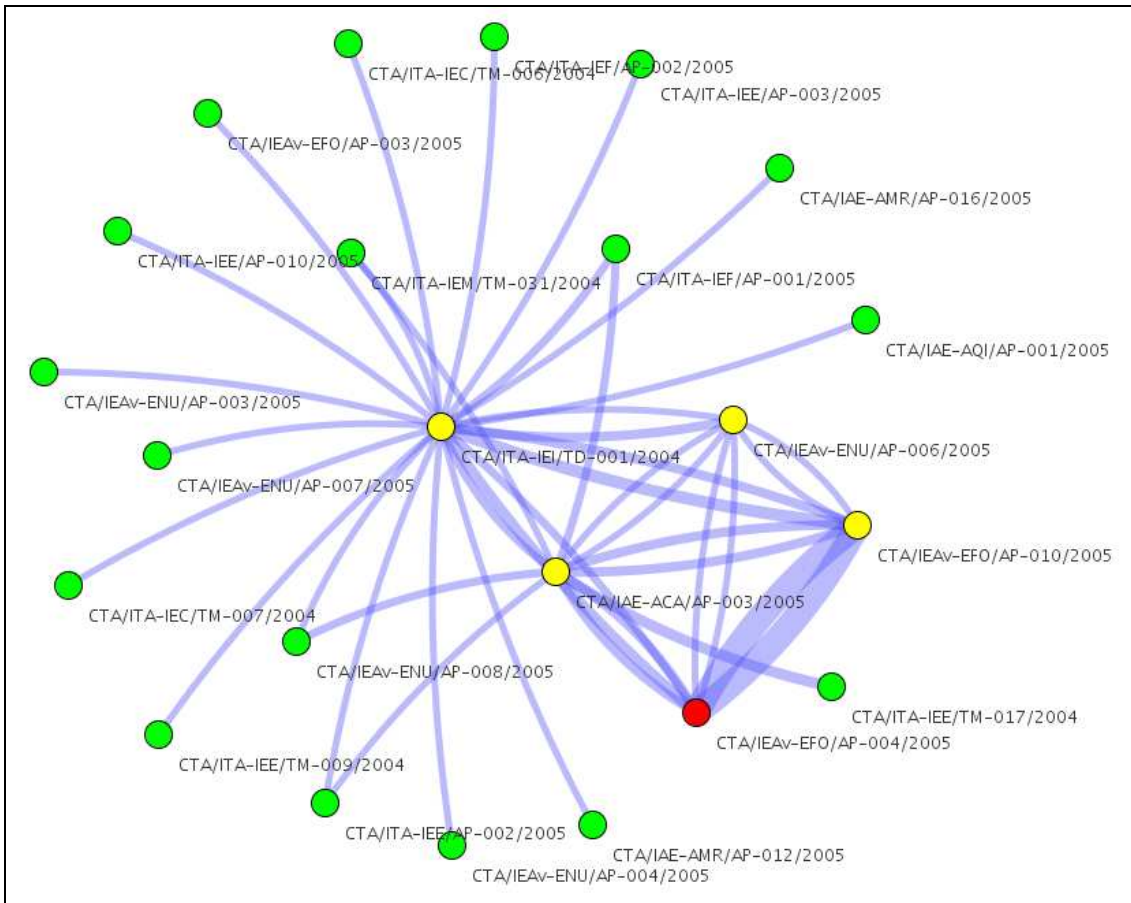


Figura 2.9 – Grafo Ligações com 5 ou mais pontos utilizando dois níveis de busca.

Também, é possível identificar através da Figura 2.9 um fato interessante, a publicação CTA/ITA-IEI/TD-001/2004 está ligada diretamente à quase todas as outras publicações do grafo, só não está diretamente ligada à uma delas. Isso ocorre provavelmente porque essa publicação tem muitas palavras-chaves cadastradas, são 69 no total enquanto na média cada publicação tem 22 palavras-chaves cadastradas, isso faz com que esta provavelmente tenha palavras-chaves em comum com muitas outras publicações.

3 CONCLUSÕES

Analisar resultados de forma exata nesse trabalho é uma tarefa difícil já que os resultados são mostrados de forma empírica, e não através de números. O que se pode concluir é que as palavras-chaves são as maiores responsáveis pelas pontuações que indicam as áreas de conhecimento abordadas nos trabalhos. Portanto, desde que preenchidas corretamente e com a devida atenção pelos autores, é perceptível que na maior parte dos casos os trabalhos sugeridos tenham boa dose de relevância para quem utiliza o sistema e que, para esse caso em especial, a solução se mostrou altamente eficiente. Contudo, se não tivéssemos as palavras-chaves bem determinadas e selecionadas segundo a análise de uma pessoa confiável provavelmente essa solução não traria resultados tão satisfatórios. Na ausência desse recurso, mantendo opção de trabalhar sobre análise de conteúdo, provavelmente as palavras-chaves teriam que ser mineradas dos títulos e resumos, e selecionadas utilizando um algoritmo de *stemming* para eliminar variações de uma mesma palavra. Isto seria uma etapa adicional relativamente trabalhosa no pré-processamento e que possivelmente não traria tão bons resultados. Neste sentido, pode se observar que a coleta da informação e o pré-processamento adequado são indispensáveis para a aplicação da mineração de dados.

É importante citar que outros métodos poderiam ser utilizados para o mesmo propósito do sistema de recomendação, e tais soluções podem ser divididas em dois grandes grupos: métodos que requerem a intervenção do usuário, e métodos que agem sem a interferência direta do usuário.

Dentre os que requerem ação do usuário, poderia ser exigido um cadastro no sistema (com nome de usuário e senha), e através do armazenamento do caminho que o mesmo percorre entre as publicações seria possível montar um perfil desse usuário e sugerir por rankear as páginas de acordo com suas preferências (WANG et al., 2008) e (SMYTH, 2007). Outra abordagem seria

fornecer ao usuário um formulário com caixas de seleção contendo palavras-chaves ou áreas de conhecimento, então o preenchimento das mesmas indicaria suas preferências (PAZZANI; BILLSUS, 2007). Além disso, também poderia ser incorporado em cada página de publicação um pequeno formulário onde o usuário avalia a mesma dizendo se recomenda ou não sua leitura e se o assunto é de interesse (PAZZANI; BILLSUS, 2007). Métodos como estes vem sendo usado pela internet com muita eficácia, mas apresentam a inconveniência de exigir um trabalho extra do usuário que, além de muitas vezes não estar disposto ceder, pode não fornecer informações confiáveis sempre.

Nesse caso, um método que funciona de forma autônoma seria uma boa opção como, por exemplo, a recomendação através da análise do conteúdo que é feita através busca de palavras-chaves (JEONG; CHO, 2004) e (PAZZANI; BILLSUS, 2007), ou selecionar as recomendações usando redes neurais (KIM et al., 2005) e (PAZZANI; BILLSUS, 2007).

Além das opções citadas, um sistema híbrido entre duas ou mais técnicas também poderia combinar, por exemplo, um sistema que dá a opção do usuário fornecer um *feedback* próprio mesclado com algum método de análise de conteúdo (BURKE, 2007).

REFERÊNCIAS BIBLIOGRÁFICAS

BURKE, R. Hybrid web recommender systems. In BRUSILOVSKY, P.; KOBASA, A.; NEJDL, W. **The adaptive web: methods and strategies of web personalization.**, Lecture Notes in Computer Science, Springer, Berlin-Heidelberg, 2007. v. 4321, p. 377-408. ISBN (978-3-540-72078-2).

DU, N.; WANG, B.; WU, B. Community detection in complex networks. In: **Journal of computer science and technology.** July 2008. p. 672-683.

JEONG, O.; CHO, D. A personalized recommendation agent system for e-mail document classification. In: **ICCSA International conference on computational science and its applications.**, 2004, May 14-17, Assisi, Italy. Proceedings... Assisi: Springer Verlag, 2004. Part III, v. 3045. ISBN 3-540-22057-7. Lecture Notes in Computer Science.

KIM, M. W.; KIM, E. J.; RYU, J, W. Collaborative filtering for recommendation using neural networks. In: **ICCSA International conference on computational science and its applications.**, 2005, May 2, Assisi, Italy. Proceedings... Assisi: Springer Verlag, 2005. v. 3480. p. 127-136. ISBN 978-3-540-25860-5. Lecture Notes in Computer Science.

PAZZANI, M. J.; BILLSUS, D. Content-based recommendation systems. In BRUSILOVSKY, P.; KOBASA, A.; NEJDL, W. **The adaptive web: methods and strategies of web personalization.**, Lecture Notes in Computer Science, Springer, Berlin-Heidelberg, 2007. v. 4321, p. 325-341. ISBN (978-3-540-72078-2).

SMYTH, B. Case-based recommendation. In BRUSILOVSKY, P.; KOBASA, A.; NEJDL, W. **The adaptive web: methods and strategies of web personalization.**, Lecture Notes in Computer Science, Springer, Berlin-Heidelberg, 2007. p. 342-376.

WANG, X.; YUAN, F.; QI, L. Recommendation in educational portal by relation based importance ranking. In: **ICWL International conference on advances in web-based learning.**, 2008, Jinhua, China. Proceedings... Jinhua: Spring Verlag, 2008. v. 5145, p. 39-48. ISBN (978-3-540-85032-8).