



**XXIII SNPTEE
SEMINÁRIO NACIONAL
DE PRODUÇÃO E
TRANSMISSÃO DE
ENERGIA ELÉTRICA**

FI/GET/06
18 a 21 de Outubro de 2015
Foz do Iguaçu - PR

GRUPO - XIV

GRUPO DE ESTUDO DE EFICIÊNCIA ENERGÉTICA E GESTÃO DA TECNOLOGIA, DA INOVAÇÃO E DA EDUCAÇÃO - GET

REDATOR DE PATENTES: FERRAMENTA PARA REALIZAÇÃO DE BUSCAS DE ANTERIORIDADE, PROSPECÇÃO DE TECNOLOGIAS, APOIO À REDAÇÃO DE PATENTES - UTILIZAÇÃO DE BASES DE DADOS DINÂMICAS

LARANJEIRA, W.S.(1);DOERNER, F.P.(1);SILVA, R.P.D.(1);JUNQUEIRA, K.G.(1);FURTADO, J.E.D.M.P.(2);LAPOLLA, L.D.C.B.(2);CAVALCANTE, G.(2); ELN(1);ELABORA(2);

RESUMO

Atualmente as empresas do setor Energético carecem de ferramentas que permitam a busca eficiente e rápida interpretação de bases brasileiras de cunho científico-tecnológico. Este tipo de informação é fundamental para a boa gestão do ciclo de vida de projetos de PD&I e da propriedade intelectual resultante. Com esse problema em vista, foi criada uma ferramenta para realização de buscas de anterioridade, prospecção de tecnologias, apoio à redação de patentes e gestão do ciclo de vida de patentes. Neste trabalho demonstramos que é possível agregar novas bases de dados à ferramenta criada, usando como exemplo a construção de uma base de dados dos anais do próprio SNPTEE.

PALAVRAS-CHAVE

data mining; análise de texto; processamento de linguagem natural; extração de palavras-chave; aprendizado de máquina; tratamento de dados;

1.0 - INTRODUÇÃO

Para inovar é preciso ter conhecer o estado da arte e monitorar sua evolução. A grande quantidade de informações a serem verificadas e o acesso a elas, em diferentes fontes e locais, são dois fatores que dificultam as atividades de busca e o acompanhamento de tecnologias. Os núcleos de inovação tecnológica (NITs) lidam cotidianamente com uma diversidade de temas e tecnologias, que competem entre si e limitam a quantidade de recursos que podem ser empregados nessas atividades. Adicionalmente, o quesito novidade é premissa básica para se obter a proteção da propriedade intelectual gerada – seja ela resultante de projetos de P&D ou soluções da engenharia – e o processo de patenteamento de uma tecnologia requer atenção a uma série de detalhes técnicos e formalismos.

Um ferramental que apóie essas atividades é de grande valor para os NITs e suas empresas, pois permite que suas equipes utilizem seus recursos intelectuais de maneira mais eficaz: na interpretação e produção de informação. Este trabalho baseia-se em uma ferramenta desenvolvida com esse conceito: a "Ferramenta de apoio a buscas de anterioridade e redação de patentes".

Tomando-a como base adicionamos uma nova fonte de informações, dos anais do SNPTEE, para realização de buscas de anterioridade e buscas prospectivas. O que permitiu, além da possibilidade de busca de anterioridade nesta base de conhecimento, o uso de diversos interpretadores estatísticos tendo em vista descrever os tópicos abordados em todos os anos do evento.

Adicionalmente, evidenciamos o valor e relevância desta base de artigos, através do processamento dos textos publicados, para geração de indicadores estatísticos dos problemas estudados em um nível de detalhamento mais granular que as linhas temáticas delimitadas pelo seminário/congresso. O esforço para o desenvolvimento dessa

ferramenta, resultou num fluxo de trabalho inovador para extração, estruturação e integração de dados científicos, de maneira rápida e eficaz.

2.0 - BUSCA DE ANTERIORIDADE

Um dos requisitos de patenteabilidade [1] é o quesito “novidade” do objeto ou invento candidato ao patenteamento. Isto se faz a partir do levantamento do estado da técnica relativo ao objeto, procurando invenções similares ou com características e propriedades similares.

Caso neste levantamento sejam encontrados objetos equivalentes, com propriedades equivalentes ou apenas com diferenças óbvias, o invento candidato não poderá ser patenteado, pois não será aprovado na fase de análise de anterioridade do Escritório de Patentes. Por outro lado, caso as propriedades dos objetos comparados possuam diferenças não-óbvias, estas precisam ser citadas, referenciadas e defendidas, de modo a demonstrar que existem diferenças sólidas, de caráter inovador e que constituem melhorias.

Esse levantamento do estado da técnica de determinado produto ou método é comumente chamado de busca de anterioridade e é feito em um acervo de patentes. Além de servir para averiguar os quesitos listados anteriormente, uma busca de anterioridade bem feita serve para evitar o dispêndio indevido de esforços e dinheiro em um pedido de patente inócuo e que certamente será negado.

A busca de anterioridade e o monitoramento da tecnologia podem prevenir conflitos, reduzir riscos de investimentos e reduzir gastos com honorários profissionais de advogados e de agentes da propriedade industrial. Além disso, a partir dos resultados da busca, o depositante terá a chance de redefinir seu invento ou projeto, dando maior atenção às características distintivas de sua inovação, à luz do que pré-existe no estado da técnica..

Costumeiramente devem ser realizadas buscas de anterioridade em diversas bases de dados, sendo mais utilizadas as bases eletrônicas e gratuitas do Instituto Nacional da Propriedade Intelectual (INPI), do United States Patent and Trademark Office (USPTO), do European Patent Office (EPO) e World Intellectual Property Office (WIPO). Os sites dessas entidades oferecem ferramentas de buscas de patentes que permitem ao inventor realizar buscas por palavras-chave, data, classificação internacional, entre outros.

Entretanto, quaisquer outras fontes de informação de cunho científico-tecnológico podem e devem ser utilizadas, pois cabem no embasamento de invalidação de um pedido de patente.

1.0 - 2.1 FONTES E BASES DE DADOS

A ferramenta desenvolvida realiza buscas em um banco de dados privado, construído a partir de informações públicas de patentes dos escritórios supramencionados, projetos da FAPESP, FAPERJ, ANEEL e FNDCT; artigos do Scielo e currículos Lattes.

Neste informe técnico demonstraremos a construção de uma base de trabalhos apresentados no SNPTEE entre os anos de 1999 e 2009.

2.2 Adicionando a base do SNPTEE

O trabalho partiu do levantamento dos artigos selecionados e apresentados no SNPTEE desde 1999. Essa base de dados foi processada e preparada para utilização pelos módulos de busca de anterioridade e apoio à análise de anterioridade e redação de patentes do software desenvolvido. O processamento envolveu as etapas de aquisição dos PDFs, extração dos respectivos textos, identificação de campos (título, autores, instituições, resumo, texto completo), consolidação dos registros em base de dados e indexação para busca.

O processo de importação de PDFs não é tão simples podendo ocorrer alguns erros. Ao final deste processo, foram adicionados como base de dados um total de 2474 artigos. A tabela abaixo mostra a disposição dos artigos por ano:

| Evento | Número de artigos |
|---------------------|-------------------|
| SNPTEE XV – 1999 | 279 |
| SNPTEE XVI – 2001 | 413 |
| SNPTEE XVII – 2003 | 335 |
| SNPTEE XVIII – 2005 | 479 |
| SNPTEE XIX – 2007 | 509 |

Tabela 1 - Distribuição dos textos importados com sucesso entre os eventos

2.3 Análise da base do SNPTEE

A base de dados de artigos do SNPTEE, estruturada, permitiu a realização da segunda etapa: a construção de um mapa de tópicos que permitiu avaliar estatisticamente os assuntos explorados ao longo dos seminários. Através da aplicação de técnicas de processamento de texto e da computação de estatísticas textuais, foi possível identificar, agregar e calcular os termos mais relevantes, segmentados por tema do simpósio. Como forma de visualização optamos pela criação de uma linha do tempo dos temas pesquisados nas edições do evento, tanto para os temas indicados pelos autores quanto para temas automaticamente identificados nos artigos.

O mapa resultante, além de explicitar o conteúdo substantivo da base de artigos, pode, por exemplo, auxiliar os organizadores do evento na criação de novos fóruns e chamadas específicas para temas atuais e de maior interesse. Com a mesma visão, é possível que empresas tenham um maior entendimento do panorama do setor, do posicionamento de seus pares e respectivos relacionamentos com determinados temas. Tendo em mãos, bases de informações locais que possam ser utilizadas em suas decisões estratégicas de P&DI. A integração de diferentes bases permite a projeção cenários e roadmaps tecnológicos para definição de cestas de projetos e determinar peso na priorização de ideias e assuntos relacionados aos projetos de P&D.

2.3.1 Agrupamentos de Tópicos - Descrição Geral

O primeiro passo do estudo consistiu no levantamento automático de algumas palavras-chave para cada documento. A ideia geral deste passo é extrair as palavras que são mais salientes no texto, posteriormente elas podem ser utilizadas para dar uma representação pequena e concisa dos documentos.

Entretanto nesta etapa estávamos apenas iniciando a prospecção da base, deste modo a representação alvo neste caso é o evento em si e não seus documentos. Após o levantamento de palavras de cada documento, foi criada uma representação da distribuição destes documentos em um espaço vetorial de N dimensões, onde cada uma das dimensões compõe uma palavra chave. Assim, cada documento foi representado por um ponto neste espaço vetorial e sua representação nos eixos tem a ver com a distribuição das palavras-chave no seu conteúdo.

A criação desta representação permitiu o uso de um algoritmo de clusterização[1] (agrupamento) que criou pequenos conjuntos de documentos de acordo com sua similaridade (com base na sua localização frente aos eixos do espaço). Ao final foram criados pequenos agrupamentos de palavras com a porcentagem de documentos pertencentes a ela. Inicialmente criamos uma disposição de 10 grupos para cada evento realizado, abaixo são mostrados somente os grupos mais significativos. Há de se lembrar que os grupos são criados automaticamente com base no espaço vetorial gerado.

SNPTEE XV - 1999 (279 documentos)

| | Documentos | Termos |
|---|------------|--|
| 1 | 34.41% | carga, temperatura, modelo, água, geração, tabela, corrente, tensão, custo, usina |
| 2 | 23.30% | transmissão, linhas, linha, custo, tensão, custos, geração, carga, fase |
| 3 | 22.58% | setor, energia elétrica, mercado, empresas, geração, desenvolvimento, custo, custos, recursos |
| 4 | 9.32% | tensão, carga, reativa, barras, barra, transmissão, potência reativa, tabela, corrente, modelo |
| 5 | 2.87% | pára, raios, pára raios, tensão, linha, descargas, sobretensões, linhas, isoladores, corrente |
| 6 | 2.51% | ambiental, gestão, custos, implantação, áreas, gestão ambiental, empresa, ações, simepar |

SNPTEE XVI - 2001 (413 documentos)

| | Documentos | Termos |
|---|------------|--|
| 1 | 46.97% | tensão, equipamentos, tabela, carga, controle, modelo, rede, custo, corrente, manutenção |
| 2 | 17.43% | mercado, setor, ambiental, empresas, desenvolvimento, custos, ambiente, serviços, atividades |

| | | |
|---|-------|--|
| 3 | 7.26% | linha, cabo, linhas, cabos, pára, raios, pára raios, corrente, tensão, circuito |
| 4 | 5.57% | carga, tensão, geração, fluxo, geradores, estabilidade, barras, ativa, barra, controle |
| 5 | 4.12% | controle, supervisão, comunicação, rede, subestações, centros, manutenção, equipamentos, real, informações |
| 6 | 3.87% | manutenção, falhas, equipamento, confiabilidade, programação, controle, disponibilidade |

SNPTEE XVII - 2003 (335 documentos)

| | Documentos | Termos |
|---|------------|---|
| 1 | 60.90% | rede, controle, custo, custos, carga, consumo, modelo, tabela, qualidade, empresa |
| 2 | 19.70% | corrente, carga, falta, controle, rede, circuito, barra, tensões, gerador, frequência |
| 3 | 7.76% | linha, linhas, cabo, cabos, modelo, tabela, isoladores, condutores, linhas transmissão, modelos |
| 4 | 3.88% | empresas, pesquisa, setor, empresa, tecnologia, gestão, setor elétrico, inovação, atividades |
| 5 | 1.79% | água, isolante, óleo mineral, papel, óleo mineral isolante, temperatura, transformadores |

SNPTEE XVIII - 2005 (479 documentos)

| | Documentos | Termos |
|---|------------|--|
| 1 | 29.65% | tensão, rede, velocidade, carga, método, vento, gerador, unidades, usina, máquina |
| 2 | 12.53% | área, ambiental, áreas, informações, gestão, região, copel, ambientais, qualidade, reservatório |
| 3 | 12.32% | rede, subestações, proteção, tensão, custos, transformador, transformadores, comunicação, custo |
| 4 | 9.81% | cabo, cabos, linhas, linha, torre, corrente, tensão, condutor, condutores, linhas transmissão |
| 5 | 9.60% | custo, mercado, usinas, consumo, período, custos, preço, carga, planejamento, demanda |
| 6 | 6.26% | proteção, falta, corrente, tensão, circuito, linha, curto, faltas, curto circuito, relé |
| 7 | 6.05% | tensão, corrente, carga, reativa, barra, fase, cargas, rede, reatores, comportamento |
| 8 | 5.85% | projetos, pesquisa, empresas, empresa, gestão, setor, conhecimento, inovação, tecnologia, recursos |

SNPTEE XIX - 2007 (509 documentos)

| | Documentos | Termos |
|---|------------|---|
| 1 | 36.15% | rede, carga, água, informações, distribuição, corrente, período, manutenção, custo, estado |
| 2 | 23.18% | corrente, manutenção, linhas, transformadores, linha, circuito, campo, fase, transformador, falha |
| 3 | 9.23% | setor, empresas, energética, avaliação, mercado, consumo, eficiência, ambiental, impactos, recursos |
| 4 | 8.64% | frequência, velocidade, rotor, gerador, turbina, unidades, unidade, máquina, resposta, excitação |
| 5 | 5.11% | usinas, mercado, preço, custo, risco, livre, usina, custos, contratos, comercialização |
| 6 | 4.13% | projetos, gestão, conhecimento, inovação, pesquisa, empresas, tecnologia, processos, recursos |

SNPTEE XX - 2009 (459 documentos)

| | Documentos | Termos |
|---|------------|--|
| 1 | 56.86% | carga, transformadores, medição, controle, campo, rede, corrente, manutenção, programa, linhas |

| | | |
|---|--------|---|
| 2 | 14.38% | custo, usinas, custos, mercado, planejamento, setor, expansão, demanda, prazo, período |
| 3 | 5.88% | proteção, linha, falta, religamento, linhas, faltas, arco, corrente, relés, relé |
| 4 | 5.01% | rede, redes, segurança, acesso, comunicação, informações, distribuição, tecnologia, tarifas, proteção |
| 5 | 3.92% | corrente, controle, circuito, curto, curto circuito, power, rede, carga, transformadores, correntes |
| 6 | 3.49% | projetos, empresas, gestão, conhecimento, inovação, setor, pesquisa, processos, indicadores |

Dentre a tabelas pode-se perceber a sobreposição de alguns termos entre diversos grupos, isso remete a interpretação de que estes termos são mais gerais e estão agregados a descrição de algum tópico menos generalista como: curto-circuito ou para-raios. Os grupos produzidos possibilitam o leitor a conclusão de diversos aspectos sobre o conteúdo das publicações, em especial pode-se observar o crescimento de publicações observando as questões de custos; ambiental; manutenção; proteção e prevenção de falhas; Observa-se que no início do evento havia mais publicações relacionadas a geração de energia, corrente que perdeu força ao longo dos anos e voltou a aparecer em 2009, relacionando-se com planejamento e expansão.

Apesar de interessantes estas conclusões são subjetivas ao analista, as relações entre diferentes tópicos não permitem provar ou inferir conclusões concretas, são somente dados que possibilitam ao leitor ter insights sobre a área ou criar algum tipo de relação com outras conclusões próprias (oriundas de outras leituras ou conhecimento prévio). Todavia análises mais granulares como serão vistas a seguir permitem a observação da evolução de um tópico em específico, o que talvez seja mais interessante do ponto de vista estratégico do evento ou do mercado.

2.3.2 *Visões Granulares Sobre Temas*

A partir do levantamento de tópicos foi possível acompanhar a evolução dos mesmos ao longo dos anos, em cada um dos eventos do SNPTEE. Foi realizada a análise baseada na porcentagem de documentos que citam o termo. Deste modo pode-se acompanhar a tendência de alta ou baixa dado um tópico em específico.

Automaticamente levantamos alguns tópicos que tiveram maior variação ao longo dos anos e com alguma observação interessante, serão mostrados a seguir. Nos gráficos a seguir observa-se, como eixo X o ano do evento, como eixo Y é observada a porcentagem de documentos do evento-ano que possuem o termo. A linha em vermelho identifica a função criada a partir da regressão linear entre os pontos, o que permite identificar a tendência simplificada do termo.

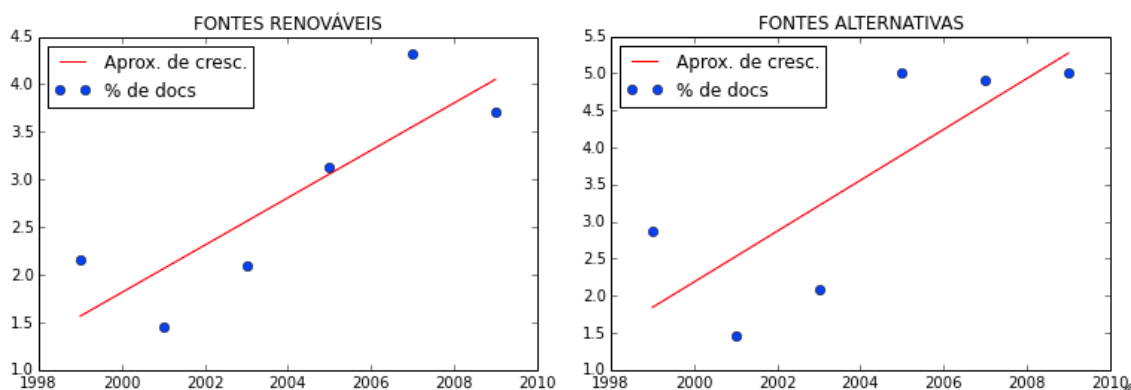


Figura 1 - Evolução da porcentagem de documentos tratando de fontes alternativas de energia.

No gráfico acima (Figura 1), podemos observar que os termos "fontes renováveis" e "fontes alternativas" tiveram tendência de alta no seu grau de importância dentro do evento. Chegando a picos de cerca de 5% dos últimos 2 eventos observados tocaram de algum modo no assunto. Assim conclui-se que, o conjunto de documentos tocantes a novos tipos de fontes de energia ou assuntos relacionados praticamente dobrou nos últimos eventos. Infelizmente não conseguimos obter os dados de 2011 e 2013 para observar a continuidade ou não desta tendência.

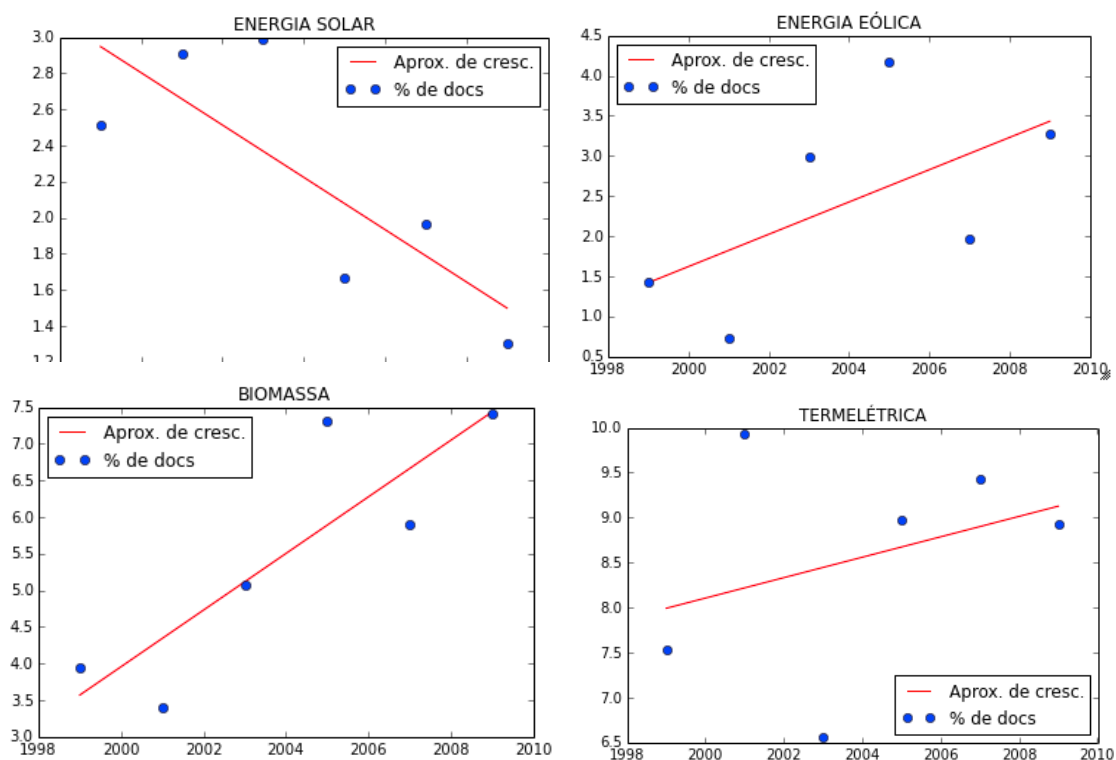


Figura 2 - Conjunto de gráficos mostrando a evolução de termos relacionados a fontes alternativas e/ou renováveis.

Na Figura 2, notamos que apesar de fontes alternativas e/ou renováveis serem itens de crescente interesse, no destaque das tecnologias inerentes a este tema, podemos perceber uma tendência de queda no âmbito de energia solar. Aqui podemos perceber o potencial deste tipo de análise, que mostra uma percepção geral sobre um tema mas também abre a possibilidade de mergulhar profundamente no tema e concluir mais sobre o assunto. Mais uma vez, a conclusão ou correlação de informações como esta é subjetiva ao analista, o que caracteriza a esta ferramenta o característica de auxílio a decisões estratégicas.

Na gráficos abaixo (FIGURA 3), apesar de óbvio, também destacamos o crescente interesse em publicações que tangem sobre a questão do meio ambiente. Porém uma observação é importante, as preocupações ambientais já apresentavam ser alvo de grande parte das publicações já em meados do ano 2000. Porém é importante ressaltar o início da preocupação socioambiental nos artigos.

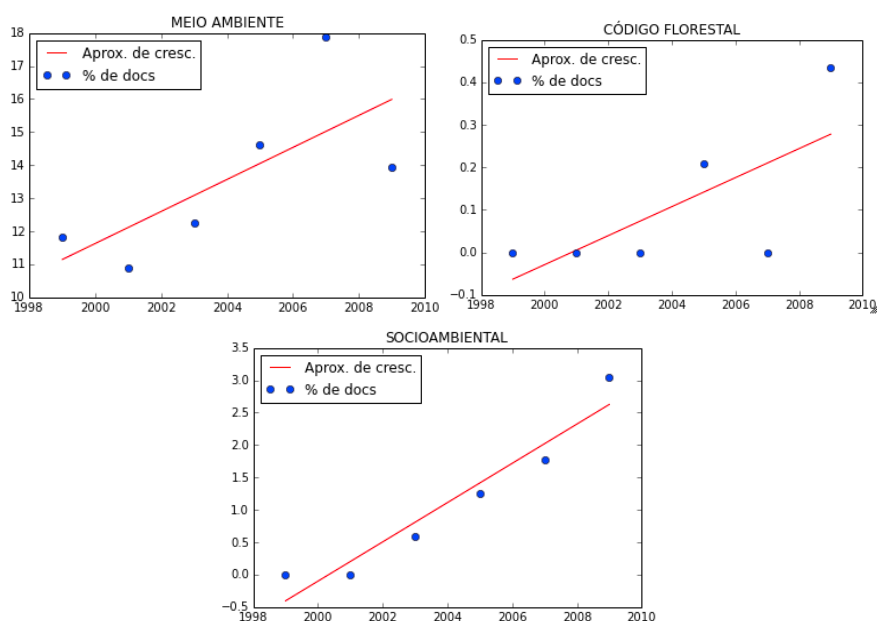


Figura 3 - Evolução dos tópicos relacionados a questão ambiental.

Vê-se que em nos meados do ano 2000 nenhum artigo abordou este tema em 2009 3% do total dos artigos abordaram esta questão, mostrando um crescimento agressivo da preocupação com a questão nos últimos eventos.

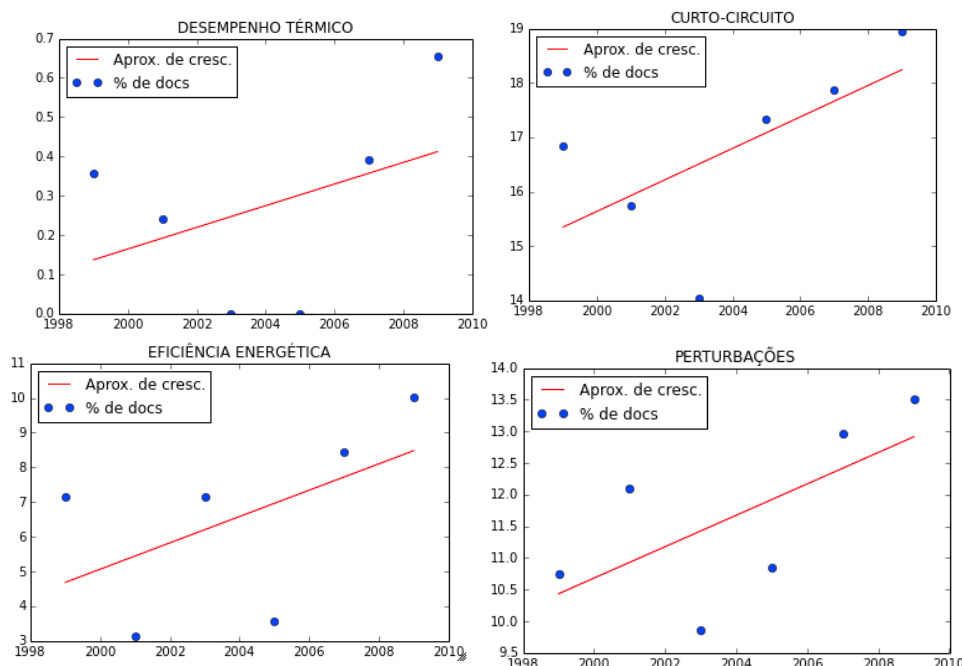


Figura 4 - Crescimento do interesse em publicações sobre qualidade e eficiência dentro do setor elétrico.

É possível observar também, o crescimento dos artigos abordado a eficiência e qualidade, tanto na geração quanto na transmissão de energia. Outros tópicos relacionados também tiveram curva de crescimento ascendente como: assoreamento, ressonância, medições fasoriais, gestão de risco, descargas, medições em campo, medição de perdas e outros.

3.0 - CONCLUSÃO

Este estudo teve por finalidade demonstrar como é possível incluir novas bases de dados no aplicativo desenvolvido para auxílio à redação de patentes e análise de anterioridade. Além de garantir maior abrangência e qualidade as buscas relativas à redação da patente, a inclusão de novas bases garante também a possibilidade de análises textuais mais complexas, assim como mostrar rotas e tendências temáticas que mais se destacaram ao longo do evento.

Deste modo, a ferramenta desenvolvida presta auxílio não somente aos núcleos de inovação presentes em grandes empresas e aos escritórios de patentes, mas também para as equipes de estratégia empresarial e inteligência competitiva. Fato este porém, que, poderá ser impulsionado pelo desenvolvimento de ferramentas de análise adequadas e customizadas a cada cenário.

A quantidade de informação levantada é grande, o que impossibilita a demonstração de todas elas no artigo, entretanto foram destacadas algumas das mais interessantes. Os dados podem ser posteriormente disponibilizados para análise. Acredita-se que as possibilidades levantadas por estas ferramentas de análise possam servir de guia em novos rumos do evento e das empresas do setor elétrico, conferindo um novo potencial e diferencial mercadológico aos interessados que fizerem uso da mesma.

4.0 - REFERÊNCIAS BIBLIOGRÁFICAS

[1] Hartigan, John A., and Manchek A. Wong. "Algorithm AS 136: A k-means clustering algorithm." *Applied statistics* (1979): 100-108.

[2] Han, Jiawei; Kamber, Micheline. "Data Mining – Concepts and Techniques" Morgan Kaufmann Publishers (2000);

5.0 - DADOS BIOGRÁFICOS

LARANJEIRA, W.S.(1);



Wilson Santana Laranjeira, Administrador, nascido em Brasília-DF, com Especialização em Gestão Estratégica da Inovação Tecnológica- Unicamp/Eletronorte (2011), formado em Administração de Empresas pela Universidade de Brasília – UnB (2003). Atualmente, trabalha na Gerência dos Programas de Pesquisa e Desenvolvimento da Eletronorte.

DOERNER, F.P.(1);



Frederico Rodolfo Parente Doerner nasceu em 1984 em Brasília - DF, é engenheiro eletricista, graduado pela Universidade Paulista – UNIP, pós-graduado em Gestão Estratégica em Inovação Tecnológica. Atualmente é engenheiro de manutenção eletrônica na Eletronorte, atuando como analista de inovação e especialista em redação de patente.