

PERFIL DA PESQUISA NA UNIPAMPA: UMA ANÁLISE BASEADA EM PALAVRAS-CHAVE DOS PROJETOS DE PESQUISA

Wolgan Ens Quepfert¹

Andrea Sabedra Bordin²

Resumo:

Dentre a gama de sistemas internos de gestão da Universidade Federal do Pampa (UNIPAMPA), encontra-se a plataforma SIPPEE (Sistema de Informação de Projetos de Pesquisa, Ensino e Extensão). Esta plataforma mantém dados de projetos de pesquisa, ensino e extensão submetidos por pesquisadores de todos os câmpus da instituição. Sistemas de informação gerenciam informações destinadas a apoiar a tomada de decisões. Contudo, manter uma grande quantidade de dados de projetos armazenados, cria a falsa sensação de que as organizações estão bem informadas à respeito do conhecimento produzido. Mais do que manter esses dados armazenados, é preciso utilizá-los de forma adequada para que os mesmos possam de fato influenciar positivamente nas decisões e estratégias das organizações. Tanto para uma gestão otimizada e racional de recursos, quanto para desenvolver estratégias para o cumprimento de metas, organizações de Ciência e Tecnologia (C&T) precisam construir mecanismos para medir as atividades de Pesquisa e Desenvolvimento (P&D). Assim sendo, o mapeamento dos projetos e produções de pesquisa da instituição pode ser utilizado como um componente auxiliar para a avaliação e medição de atividades científicas e tecnológicas. As palavras-chave de um projeto de pesquisa, têm como objetivo transmitir ao leitor de uma forma sucinta a temática do trabalho e também auxiliar ferramentas indexadoras e motores de busca à encontrar a publicação. Desta forma, o presente trabalho propõe uma análise do perfil de pesquisa da UNIPAMPA, através da extração e análise das palavras-chave dos projetos de pesquisa da plataforma SIPPEE. A execução da atividade proposta, está dividida em quatro etapas principais: (I) conversão dos arquivos PDF dos projetos em arquivos de texto pleno; (II) extração das palavras-chave; (III) criação dos perfis dos pesquisadores; (IV) armazenamento das informações em um banco de dados NoSQL. Com a construção desta base de dados, cria-se a possibilidade de realizar diversos tipos de consultas. A fim de realizar uma análise e exibir o perfil temático e os tópicos mais abordados nos projetos de pesquisa, foram definidas as seguintes consultas: palavras-chave mais utilizadas; pesquisadores de um determinado tópico (palavra-chave); número de projetos por autor; tópicos de pesquisa (perfil) de um pesquisador. Através da análise das consultas podemos verificar quais os termos mais frequentemente utilizados, podendo avaliar se a pesquisa está muito centralizada, esparsa ou bem segmentada nos tópicos de pesquisa existentes. O número de projetos submetidos por pesquisador permite verificar numa análise prévia os pesquisadores mais ativos da instituição. Os termos mais utilizados pelos pesquisadores, mostram quais as principais áreas de interesse de pesquisa,

proporcionando uma visão de onde está sendo empregado o capital intelectual durante a criação de conhecimento. Sistemas de Ciência e Tecnologia (C&T), mais especificamente os responsáveis por armazenar informações a respeito do desenvolvimento da pesquisa e produção da inovação de uma instituição, devem ser utilizados como ferramenta crítica no suporte a tomada de decisões por parte da gestão, pois possuem informações implícitas de grande valor sobre as atividades de pesquisa que são exercidas por seus pesquisadores.

Palavras-chave: Extração de Informação, Sistemas de Apoio à Decisão, Perfil de Pesquisa, UNIPAMPA

Modalidade de Participação: Iniciação Científica

PERFIL DA PESQUISA NA UNIPAMPA: UMA ANÁLISE BASEADA EM PALAVRAS-CHAVE DOS PROJETOS DE PESQUISA

¹ Aluno de graduação. ens.wolgan@gmail.com. Autor principal

² Docente. andreabordin@unipampa.edu.br. Orientador



PERFIL DA PESQUISA NA UNIPAMPA: UMA ANÁLISE BASEADA EM PALAVRAS-CHAVE DOS PROJETOS DE PESQUISA

1. INTRODUÇÃO

Dentro de sua gama de sistemas internos de gestão da Universidade Federal do Pampa (UNIPAMPA), encontra-se a plataforma SIPPEE (Sistema de Informação de Projetos de Pesquisa, Ensino e Extensão). A função desta plataforma é de gerenciar projetos de pesquisa, ensino e extensão submetidos por pesquisadores de todos os câmpus da instituição.

Sistemas de informação são conjuntos de componentes inter-relacionados que gerenciam informações destinadas a apoiar a tomada de decisões (LAUDON & LAUDON, 2007). Contudo, manter uma grande quantidade de dados de projetos armazenados, cria a falsa sensação de que as organizações estão bem informadas à respeito do conhecimento produzido (CARDOSO; MACHADO, 2007). Mais do que manter esses dados armazenados, é preciso utilizá-los de forma adequada para que os mesmos possam de fato influenciar positivamente nas decisões e estratégias das organizações.

Segundo Hayashi et al. (2006) tanto para uma gestão otimizada e racional de recursos, quanto para desenvolver estratégias para o cumprimento de metas, organizações de Ciência e Tecnologia (C&T) precisam construir mecanismos para medir as atividade de Pesquisa e Desenvolvimento (P&D). Assim sendo, o mapeamento dos projetos e produções da instituição pode ser utilizado como um componente auxiliar para a avaliação e medição de atividades científicas e tecnológicas.

Os projetos cadastrados na plataforma SIPPEE têm grande potencial de utilização e podem ser explorados, por exemplo, a fim de analisar o perfil de pesquisa da UNIPAMPA ajudando na tomada de decisões por parte da gestão.

Uma das atividades na produção de um trabalho de pesquisa científica, mais especificamente durante sua escrita, é a escolha das palavras-chave. As palavras-chave tem como objetivo transmitir ao leitor de uma forma sucinta a temática do trabalho e também auxiliar ferramentas indexadoras e motores de busca à encontrar a publicação, colaborando em aspectos como alcance e citações do trabalho.

Desta forma, o presente trabalho propõe uma análise do perfil de pesquisa da UNIPAMPA, através da extração e análise das palavras-chave vinculadas aos projetos de pesquisa da plataforma SIPPEE.

2. METODOLOGIA

A execução da atividade proposta, está dividida em quatro etapas principais: (I) conversão dos arquivos *PDF* dos projetos em arquivos de texto pleno; (II) extração das palavras-chave; (III) criação dos perfis dos pesquisadores; (IV) armazenamento das informações. A Figura 1 proporciona uma visão geral da execução destas etapas.

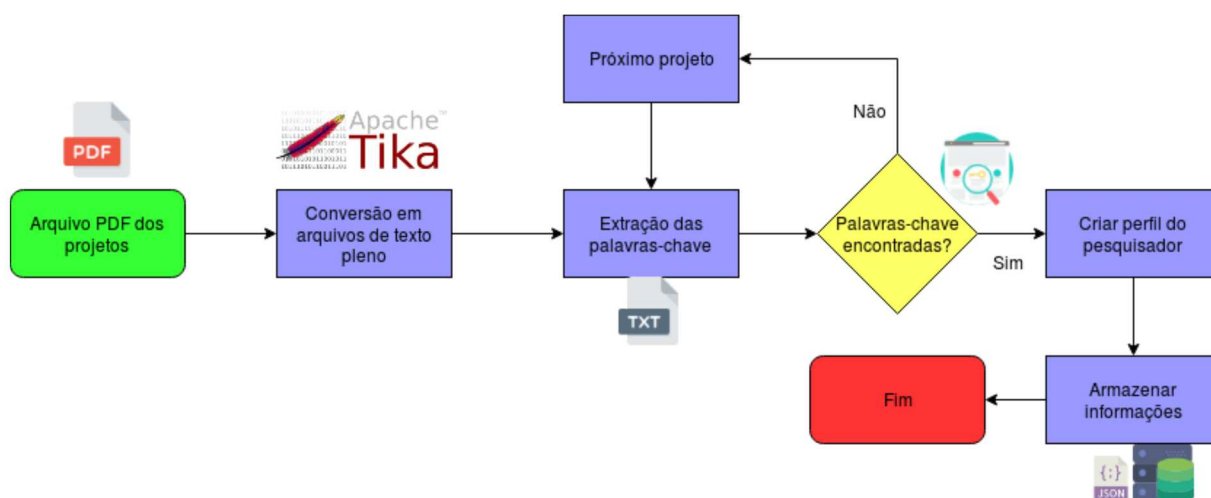


Figura 1: Extração e armazenamento das palavras-chave e armazenamento dos perfis dos pesquisadores e seus projetos de pesquisa.

O primeiro passo a ser realizado consiste na conversão dos arquivos *PDF* dos projetos de pesquisa cadastrados e aprovados na plataforma SIPPEE em arquivos de texto pleno. A Tabela 1 apresenta o número de projetos de pesquisa cadastrados na plataforma SIPPEE oriundos dos diversos campi da UNIPAMPA.

Tabela 1: Número de projetos de pesquisa cadastrados na plataforma SIPPEE

Campus	# Projetos
Alegrete	210
Bagé	259
Caçapava do Sul	88
Dom Pedrito	165
Itaqui	248
Jaguarão	91
Santana do Livramento	117
São Borja	187
São Gabriel	134
Uruguaiana	700
Total	2199

Essa conversão é necessária devido a maior facilidade de se manipular arquivos de texto em comparação à manipulação de arquivos *PDF*, sendo utilizada a ferramenta *Apache Tika* para fazer a conversão dos arquivos. Esta escolha se deve

ao fato da ferramenta ter apresentado arquivos de texto com uma qualidade (estrutura do texto) superior a outras ferramentas.

Na segunda etapa, já com os arquivos de texto dos projetos em mãos, foram extraídas as palavras-chave através do uso de expressões regulares da linguagem de programação *Python*. Durante este processo, foram detectados quatro templates distintos de arquivos de projeto, acarretando na elaboração de diversas expressões regulares tais como as exibidas na Figura 2.

```
templates = [  
    {  
        "template_re": "Palavras-chave(.*\s*.*)EQUIPE",  
        "kw_start_line_re": "Palavras-chave",  
        "kw_end_line_re": "EQUIPE",  
        "jmp_first_line": False  
    },  
    {  
        "template_re": r"5.\s*PALAVRAS-CHAVE",  
        "kw_start_line_re": "PALAVRAS-CHAVE",  
        "kw_end_line_re": "PERÍO",  
        "jmp_first_line": False  
    }  
]
```

Figura 2: Expressões regulares para detecção dos diversos templates de projetos.

Comumente as palavras-chave possuem uma “seção” exclusiva no texto, projetos onde as palavras-chave não estavam presentes no texto foram ignorados nesta etapa.

Após extrair as palavras-chave dos arquivos de texto, foi construído um perfil de pesquisador contendo seu nome, câmpus e palavras-chave. O perfil é armazenado em uma base de dados NoSQL (MongoDB).

Com a construção desta base de dados, cria-se a possibilidade de realizar diversos tipos de consultas. A fim de realizar uma análise e demonstrar o perfil temático e os tópicos mais abordados nos projetos de pesquisa, foram definidas as seguintes consultas: palavras-chave mais utilizadas; pesquisadores de um determinado tópico(palavra-chave); número de projetos por autor; tópicos de pesquisa (perfil) de um pesquisador. O número de registros por consulta foi definido em oito, de forma arbitrária.

3. RESULTADOS e DISCUSSÃO

Nesta seção são apresentados e discutidos os resultados referentes às consultas definidas anteriormente. Para cada consulta um gráfico é apresentado a fim de facilitar a visualização das informações geradas e realizada uma breve discussão sobre a relevância e potencial de utilização dos dados.

A Figura 3 apresenta dados referentes às palavras-chave que mais foram utilizadas nos projetos de pesquisa, onde podemos observar que o termo “enfermagem” é o termo mais frequentemente utilizado. Essas informações podem

ser utilizadas para verificar por exemplo, se a pesquisa está muito centralizada, esparsa ou bem segmentada nos tópicos de pesquisa existentes.

Na Figura 4 são exibidos os pesquisadores que submeteram trabalhos que incluem “enfermagem” como palavras-chave e o número de ocorrências dessa palavra em seus trabalhos, dessa forma, pode-se verificar quais pesquisadores fazem parte dos principais tópicos de pesquisa atualmente na instituição, além do conhecimento de quais pesquisadores trabalham em linhas de pesquisa semelhantes.

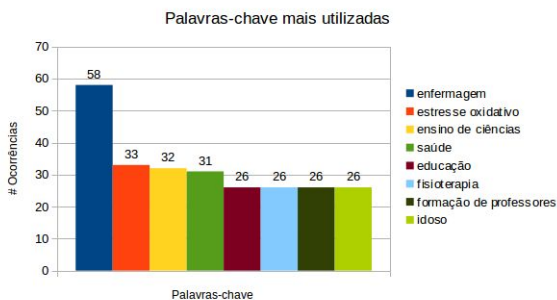


Figura 3: Palavras-chave mais utilizadas.

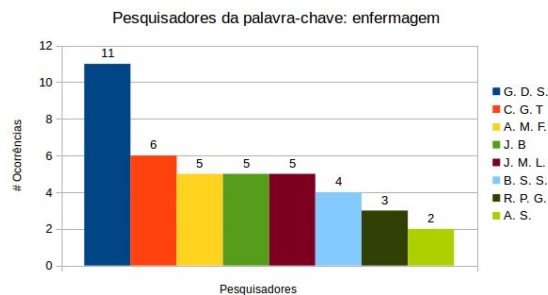


Figura 4: Pesquisadores da palavra-chave enfermagem.

A Figura 5 mostra dados referentes ao número de projetos de pesquisadores com mais projetos submetidos à plataforma SIPPEE, onde podemos verificar que o pesquisador G. G. B possui o maior número de projetos (28) na plataforma. Com essas informações, é possível verificar numa análise prévia (sem considerar outros aspectos como por exemplo citações e qualificação dos trabalhos) os pesquisadores mais produtivos.

Na Figura 6, é exibido o perfil do pesquisador G. G. B., ou seja, quais são os termos (palavras-chave) mais utilizados por ele, podendo assim verificar quais as principais áreas de atuação de cada pesquisador da instituição, proporcionando uma visão de onde está sendo empregado o capital intelectual dos pesquisadores no processo de criação de conhecimento.

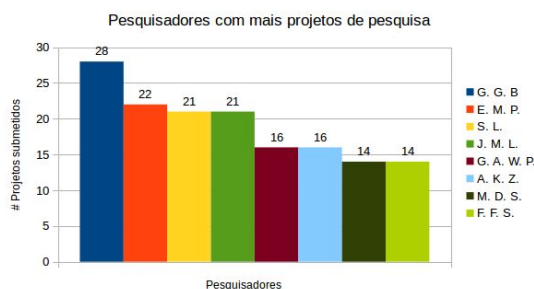


Figura 5: Número de projetos por pesquisador



Figura 6: Total de projetos por pesquisador

4. CONSIDERAÇÕES FINAIS

Sistemas de Ciência e Tecnologia (C&T), mais especificamente os responsáveis por armazenar informações a respeito do desenvolvimento da pesquisa e produção da inovação de uma instituição, devem ser utilizados como ferramenta crítica no suporte a tomada de decisões por parte da gestão, pois possuem informações implícitas de grande valor sobre as atividades de pesquisa que são exercidas por seus pesquisadores.

Analisando os dados extraídos, podemos observar quais são os tópicos de pesquisa mais pesquisados pelos pesquisadores da instituição, o perfil desses pesquisadores e como isso se relaciona com a produtividade (no que diz respeito a elaboração de projetos de pesquisa) de cada câmpus da UNIPAMPA.

Torna-se cada vez mais importante a padronização no processo de cadastramento de projetos de pesquisa nos sistemas de gestão de conhecimento das instituições, com o intuito de facilitar a construção e funcionamento de ferramentas capazes de extrair informações tão valiosas na construção de estratégias e tomada de decisão da gestão de pesquisa.

Este trabalho é parte de um trabalho futuro que visa recomendar pesquisadores com potencial de colaboração em projetos de pesquisa, visando fomentar a colaboração científica na instituição.

5. REFERÊNCIAS

Apache Tika - a content analysis toolkit [internet]. [acesso em 27 ago 2017]. Disponível em: <https://tika.apache.org/>.

CARDOSO, Olinda NP; MACHADO, Rosa TM. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. Revista de administração pública, v. 42, n. 3, p. 495-528, 2008.

HAYASHI, Maria Cristina Piumbato Innocentini et al. Indicadores de CT&I no Pólo Tecnológico de São Carlos: primeiras aproximações. RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação, v. 3, n. 2, p. 17-30, 2006.

LAUDON, Kenneth C; LAUDON, Jane P. Sistemas de Informações Gerenciais. 7. ed. São Paulo: Pearson Prentice Hall, 2007.

MongoDB - MongoDB for GIANT Ideas [internet]. [acesso em 15 ago 2017]. Disponível em: <https://www.mongodb.com/>.