

PIBIFSP	PROJETO DE PESQUISA
---------	---------------------

TÍTULO DO PROJETO: Desenvolvimento de uma ferramenta Web Analytics.

Área do Conhecimento (Tabela do CNPq): 1 . 0 3 . 0 3 . 0 4 - 9

1. RESUMO

Web Analytics (análise de dados da *web*) é um tema muito presente nos dias de hoje, pois envolve técnicas de tratamento de grande volume de dados (massa de dados) visando a detecção de padrões de comportamento e tendências, o que auxilia na tomada de decisões em um negócio ou organização. Este projeto tem como foco pesquisar os princípios relacionados à análise de dados coletados no âmbito da web e desenvolver uma ferramenta que implemente um ou mais métodos escolhidos, apresentando resultados, a partir dos dados obtidos, em relatórios gerados pela própria ferramenta. Nesse sentido, pretende-se desenvolver uma ferramenta Web Analytics que colete dados – no lado *server* (*server-side*) ou no lado *client* (*client-side*) – a partir dos acessos às páginas localizadas em servidores *web* organizacionais e/ou institucionais e apresente resultados de interesse dos proprietários dos *sites* e/ou servidores onde foram efetuadas as coletas de dados.

2. FUNDAMENTAÇÃO TEÓRICA

Web Analytics baseia-se em análise inteligente dos dados coletados de *sites* monitorados, sendo as coletas feitas a partir das visitas e interações realizadas pelos usuários (SANTOS JR, 2008). Essa análise pode ser de conteúdo (texto e documentos contidos nos sites), de estrutura (páginas e sites cruzados) e de uso (comportamento dos usuários), sendo este último o tipo adotado no presente trabalho.

Na etapa de coleta, os dados podem ser obtidos de registros gravados no servidor *web* (arquivos de *log*) ou via aquisição efetuada no *client-side* (scripts acrescentados ao código-fonte) da página *web*.

Posteriormente à coleta, é feita a mineração dos dados (*data mining*). Mineração é um processo de exploração de grande quantidade de dados (*big data*, ou massa de dados), com o objetivo de encontrar tendências e padrões comportamentais. Ela pode utilizar-se de diversas e distintas técnicas e áreas, como Aprendizado de Máquina (*Machine Learning*), *Data Warehousing*, Banco de Dados, Estatística, Visão de Dados, etc. Nesse processo, os dados passam por um pré-processamento (limpeza, redução e seleção) e, depois, um processamento (classificação, ordenação, agrupamento e acúmulo), conforme métricas pré-estabelecidas (URLs acessadas, número de cliques sobre cada *hyperlink* em um dado período de tempo, tempo médio de permanência em cada URL acessada, etc.), com o objetivo de permitir detectar preferências e tendências dos usuários que acessam esses sites. Sob essa óptica, a mineração de dados está relacionada à Inteligência Artificial (IA), e a principal técnica utilizada aqui é o Aprendizado de Máquina, em que se procura generalizar os resultados encontrados para um nível compreensível aos seres humanos. Também são utilizados conceitos e técnicas relacionados à área

da Estatística (a fim de estudar formas de coleta, organização e interpretação dos dados numéricos, levando à descoberta de padrões por meio do cálculo de aproximações, correlações, médias, desvios e taxas de erro) e a Visualização de Dados (com o intuito de refinar os resultados obtidos, melhorando a capacidade de entendimento) (REZENDE, 2005; SAS, 2016). Bases de Dados e *Data Warehousing* são elementos auxiliares no processo, haja vista a necessidade de os dados estarem armazenados de forma estruturada em um repositório correspondente ao domínio que se deseja analisar, bem como que esse repositório tenha como suporte um ambiente físico eficiente e confiável, o que demanda um estilo de gerenciamento que remete a *Data Warehouses*. Cabe observar que *Data Warehousing* é um elemento facultativo, não sendo considerado neste trabalho.

As métricas acima referidas são o alicerce para a definição dos Indicadores Chave de Performance, ou KPIs (*Key Performance Indicators*). As KPIs serão a estrutura de todo o processo de análise, pois, a partir delas, é possível saber se e onde os investimentos de tempo e recursos nos sites apresentam o retorno esperado (CARNEIRO, 2008).

Finalmente, Web Analytics serve como auxílio para as equipes de Marketing, e pode ser integrado a sistemas de BI (*Business Intelligence*) (SANTOS JR, 2008), que poderão utilizar as informações obtidas a partir dos dados coletados para tomar decisões de negócio.

3. OBJETIVOS

3.1. OBJETIVO GERAL

Implementar uma ferramenta Web Analytics que auxilie gerentes de conteúdo de *sites* na tomada de decisões a partir da análise dos dados coletados a partir das visitas realizadas e registradas.

3.2 OBJETIVOS ESPECÍFICOS

- Definir, junto aos interessados (para o desenvolvimento da ferramenta ora proposta, os responsáveis pelo *site* do IFSP campus Cubatão), as métricas apropriadas, conforme suas necessidades particulares;
- Definir formas de mineração de dados a partir das métricas definidas;
- Definir os algoritmos de análise de dados compatíveis com as formas de mineração definidas;
- Implementar ferramenta de Web Analytics.

4. MATERIAIS E MÉTODOS

A metodologia de pesquisa se baseia em pesquisas feitas em livros, *sites* e consulta com professores e técnicos do próprio campus, sendo ainda possível busca em outros tipos de fontes (artigos, revistas, etc.). Os materiais utilizados serão os disponíveis no próprio Instituto Federal: computadores, softwares, IDEs e livros relacionados ao tema existentes na biblioteca.

Em princípio, o projeto não conta com nenhum tipo de financiamento, nem interno nem externo. A pesquisa sobre técnicas de análise será feita com professores e em livros que tratem do assunto, para que se possam definir as melhores formas de análise de dados e desenvolvimento do sistema.

Como interessados, conforme mencionado nos objetivos específicos, serão consultados, em princípio, os responsáveis pelo *site* do IFSP, campus Cubatão.

Além da concretização da implementação ora proposta, pretende-se que este projeto seja utilizado em pesquisas subsequentes, permitindo, ainda, que sejam ampliadas as técnicas estudadas durante seu desenvolvimento. Inicialmente, está prevista a utilização do paradigma de orientação a objetos, bem como *frameworks open-source*, para implementar a ferramenta. Futuramente, será possível que sejam feitos estudos mais precisos e avançados envolvendo aquisição de conhecimento, aprendizagem de máquina e, aprofundando, mineração de dados.

5. PLANO DE TRABALHO

Tabela 5.1 Metas estabelecidas para a pesquisa.

METAS	DESCRIÇÃO
1	Estudo e pesquisa sobre estatística.
2	Estudo e pesquisa sobre inteligência artificial.
3	Estudo e pesquisa sobre mineração de dados.
4	Estudo e pesquisa sobre análise de dados.
5	Relatório Parcial entrega até 15/11/18
6	Especificação e documentação dos algoritmos e métodos de análise.
7	Implementação computacional e documentação da ferramenta.
8	Elaboração do relatório final.
9	Relatório Final entrega até 29/03/2019

Tabela 5.2 Cronograma proposta para cumprimento das metas.

METAS	MESES								
	JUL	AGO	SET	OUT	NOV	DEZ	JAN	FEV	MAR
1	X	X							
2		X	X						
3			X	X					
4				X	X				
5					X				
6					X	X	X		
7						X	X	X	
8								X	
9									X

6. VIABILIDADE DE EXECUÇÃO

Este projeto será desenvolvido no campus de Cubatão do IFSP, que disponibilizará seus recursos: computadores, livros, softwares e IDEs, bem como seu espaço (salas de estudo, biblioteca e laboratórios).

Em princípio, o projeto não conta com nenhum financiamento, interno ou externo. O apoio previsto esperado é o de professores de matemática, os quais serão de grande ajuda em questões que envolvem estatísticas relacionadas ao projeto, e técnicos interessados, principalmente os responsáveis pela hospedagem e manutenção de páginas institucionais em servidores.

7. RESULTADOS ESPERADOS E DISSEMINAÇÃO

Ao final deste projeto, espera-se ter uma ferramenta Web Analytics desenvolvida, a qual contará com uma API de coleta e análise de dados, junto a uma aplicação *web* que ofereça recursos de responsividade e portabilidade.

Também será elaborada uma documentação com os métodos e as técnicas utilizadas.

Os resultados a serem obtidos servirão para fomentação e disseminação do assunto e como base para artigos científicos e participação em eventos acadêmicos internos e externos.

REFERÊNCIAS BIBLIOGRÁFICAS

CARNEIRO, R. **Web Analytics**: Uma visão brasileira. p. 13-18. Disponível em: <http://www.e-commercebrasil.org/down/web.analytics.uma.visao.brasileira.pdf>. Acesso em: 22 jun. 2018.

REZENDE, S. O. **Sistemas Inteligentes**: fundamentos e aplicações. 1. ed. Barueri: Manole, 2005.

SANTOS JR, A. **Web Analytics**: Uma visão brasileira. p. 9-12. Disponível em: <http://www.e-commercebrasil.org/down/web.analytics.uma.visao.brasileira.pdf>. Acesso em: 26 jun. 2018.

SAS - Software de Business Analytics e Business Intelligence. **Mineração de dados**: o que é e qual sua importância? Disponível em: https://www.sas.com/pt_br/insights/analytics/mineracao-de-dados.html. Acesso em: 28 jun. 2018.