

Redes de Compartilhamento de Conhecimento

Uma análise exploratória dos dados

Carlos Henrique Pacheco de Souza

30 de Abril de 2017

Universidade Federal de Minas Gerais

Motivação

Desenvolvimento de software

- Desafios que vão além do conhecimento técnico.
- Escolha de quais tecnologias e ferramentas utilizar.
- Aspectos sociais e culturais ignorados.
- Mão de obra especializada.
- Custo e sucesso do projeto, influenciados pela demanda desses profissionais.
- Distribuição do conhecimento técnico em diferentes regiões.

GitHub

Redes Sociais

 **stackoverflow**

Visão Tradicional x Objetivo

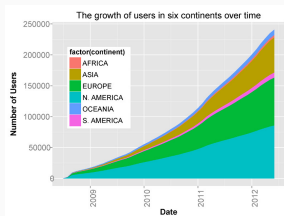


Figure 1: Taxa de crescimento do StackOverflow.

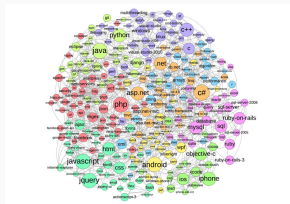


Figure 2: 256 tecnologias do StackOverflow.



Figure 3: Usuários do StackOverflow.



Figure 4: Usuários do GitHub.

Desafios

- Transforma descrições em coordenadas geográficas.
- Endereços são as principais formas de expressão da localização.
- Variações no endereço(abreviações,omissões e erros) dificultam.

Aplicação de conhecimentos obtidos em outras disciplinas:

- Mineração de Dados
- Banco de Dados Avançados(NoSQL)
- Processamento de Linguagem Natural
- Algoritmo e Estrutura de Dados(Paralelismo e Programação Dinâmica)
- Banco de Dados Geográficos

GHTorrent

StackExchange



GeoNames

mongoDB.



Etapas

Cronograma

Etapa \ Semana	07/10	18/11	15/01	21/04	26/05	23/06
Obter informações das bases de dados.	100%					
Mapear e extrair os dados pertinentes ao problema.		100%				
Realizar um pré-processamento dos dados.			100%			
Desenvolver uma interface para visualização dos dados.				80%		
Validar as informações de localizações com serviços web.					0%	
Estudar a correlação dos dados.						0%

Perguntas?