

Mestrado em Informática
Proposta de Trabalho
Extração e Análise de dados textual on line para verificar a
viabilidade de startups na área de tecnologia

Ricardo Moraes Guimarães¹

¹Universidade Federal do Amazonas - UFAM

`ricardommoraes@gmail.com`

1. Introdução

O aumento do uso e popularidade da internet nos trouxe acesso livre a vários tipos de informações de produtos, pessoas, empresas, comércios entre outros. Essas mudanças transformaram a *web* em um imenso repositório de informações para o mundo inteiro. Pesquisas relacionadas com geração de conhecimento a partir de informações extraídas dessas fontes públicas e acessíveis de dados são úteis, fascinantes e ao mesmo tempo desafiadoras [Liu 2011].

Mais recentemente houve um interesse em desenvolver técnicas de pesquisas e ferramentas para realizar pesquisas inteligentes para análise de negócios, utilizando dados textuais para gerar informações sobre negócios, marketing ou aspectos de inovação de alguns setores da indústria. As pesquisas são realizadas com dados encontrados na *web*, blogs, fóruns ou sites de review e redes sociais. As informações podem ser colhidas também a partir de informações online de determinadas empresas [Davide 2015].

Tais pesquisas podem ser feitas utilizando técnicas e métodos analíticos

Tais informações podem ser utilizadas para análises de viabilidade de startups levando em consideração fatores externos e internos referente ao empreendimento.

2. Objetivos

Neste trabalho, buscaremos métodos de extração e análise de dados textuais a partir das pesquisas realizadas na web e aplicadas ao principal componente analisado (PCA) para verificar os fatores de sucesso na criação de uma startup. É importante ressaltar os seguintes objetivos para realização do trabalho:

- Elaborar e desenvolver métricas para a análise e uso de informações, aplicadas em startup.
- Desenvolver técnica de extração de dados textual aplicado ao principal componente analisado (PCA).
- Elaborar e desenvolver métricas para a análise e uso de informações, aplicadas em startup.

3. Metodologia

4. Cronograma

Tabela 1: Cronograma

Primeiro Ano	Meses					
Atividades	1-2	3-4	5-6	7-8	9-10	11-12
Cursar Disciplinas						
Levantamento Bibliográfico						
Estudo do Problema						
Definição da Metodologia						
Defesa da Proposta de Dissertação						
Segundo Ano	Meses					
Atividades	1-2	3-4	5-6	7-8	9-10	11-12
Validação da Metodologia						
Elaboração da Dissertação						
Defesa da Dissertação						

5. Trabalhos Relacionados

Existem alguns trabalhos relacionados com a extração de dados textuais online, dentre eles, consideramos o trabalho de Giacomo di Tollo

Este trabalho possui semelhanças a abordagem utilizada no trabalho de Go e Bhayani, no entanto, fazemos abordagem na língua portuguesa com suas diversidades principalmente no campo semântico.

6. Resultados Esperados

7. Conclusão

Referências

- Davide, D. T. G. T. S. L. G. D. M. (2015). Using online textual data, principal component analysis and artificial neural networks to study business and innovation practices in technology-driven firms. *Computers in Industry*.
- Liu, B. (2011). *Web Data Mining - Exploring Hyperlinks, Contents and Usage Data*. Springer, second edition.