




THE DEVELOPER'S CONFERENCE

**Usando LDA para seleção de tópicos para Recuperação
de Informação aplicado a Inteligência de Mercado
Trilha Machine Learning**

Patricia Nunes Gonçalves



PHD Engenharia da Computação – Universidade de Lisboa
Arquitetura Corporativa  **Sicredi**

Agenda



- Contexto
- Introdução Inteligência Mercado
- Algoritmo LDA - Latent Dirichlet allocation
- Experimento
- Resultados
- Conclusão



A primeira instituição financeira cooperativa do Brasil

Em um mundo cada vez mais colaborativo, queremos unir pessoas com os mesmos propósitos. Por isso, há mais de 116 anos, acreditamos que é possível lidar com o dinheiro de uma maneira diferente.

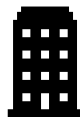
Saber mais:
www.sicredi.com.br



4 milhões de Clientes



R\$ 2,7 bilhões em resultados



R\$ 15 bilhões patrimônio líquido



THE
DEVELOPER'S
CONFERENCE

POR QUE MONITORAR O MERCADO ?

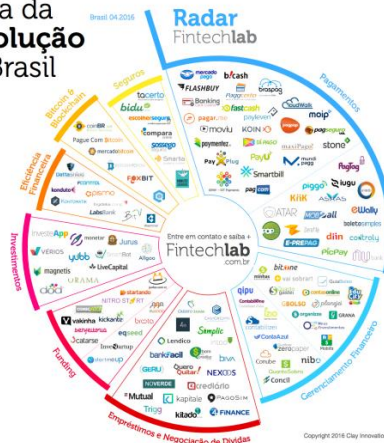
...a Nova Economia está mudando completamente o cenário concorrencial do mercado e empresas inovam para competir...



Grandes players aceleraram inovação e transformação digital



Mapa da **Revolução** no Brasil



A guerra dos bancos chega ao pedágio

Com a Veloe, Bradesco e Banco do Brasil querem enfrentar a ConectCar, controlada pelo Itaú Unibanco e pelo Grupo Ultra nos pagamentos eletrônicos

Toshiba vai vender alface no Japão

A China conquista o Brasil

Os investimentos chineses por aqui já chegam a R\$ 100 bilhões. Das commodities à tecnologia, o gigante asiático está firmando suas garras na indústria nacional, como parte de um objetivo maior: ditar as regras da globalização.

Cientes do Banco do Brasil já podem fazer transações financeiras no WhatsApp

Mercado Livre recebe aval do Banco Central para virar financeira

Zuckerberg revela novas pistas sobre o "Facebook do futuro", com pagamentos e duas empresas separadas

...a Nova Economia está mudando completamente o cenário concorrencial do mercado e empresas inovam para competir...



THE
DEVELOPER'S
CONFERENCE

O MERCADO MUDOU.

Desafio de permanecer competitivo!

Grande
inovação

PI

Clientes
transações

Mercado Livre recebe aval do Banco Central para virar financeira

Zuckerberg revela novas pistas sobre o "Facebook do futuro", com pagamentos e duas empresas separadas

Inteligência Competitiva



Modelo estruturado de levantamento e análise de dados e fatos não estruturados com o objetivo de **apoiar a tomada de decisão**.

*Estruturar olhar **para o mercado**, avaliando não só concorrentes, como o setor, macro tendências, insights, sinais fracos, para assim **gerar análises e criar hipóteses e ações**.*

Suporte para decisão

Processo sistemático, estruturado e ético

Caráter antecipativo
Macrotendências

Insumos de mercado para direcionamentos e priorizações



Informação é matéria-prima

Escopo amplo, não apenas concorrentes

Processo recorrente de revisão da estratégia



THE DEVELOPER'S CONFERENCE

Através do celular, Internet, e-mails, televisão, rádio, jornais, livros, mídias sociais, entre outras fontes, pessoas recebem diariamente 105 mil palavras (23 palavras por segundo durante o período em que estão acordadas)

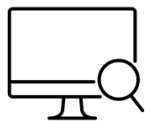


34 GB / DIA!

Fonte: Universidade da Califórnia, San Diego



Ciclo Inteligência Competitiva



Coleta



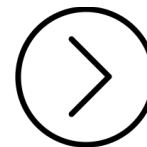
Análise



Reporte



Disseminação



Execução

Objetivos 2019 foram...



Exploração de Inteligência Artificial em Inteligência Competitiva

[1] Filtros

Como podemos usar Inteligência Artificial para aprimorar os filtros de pesquisa de notícias para o Radar de Inteligência Competitiva?

[2] Atualização

Como podemos proceder a atualização desses filtros de forma automática?

[3] Extração Relações

Através da IA realizar correlações entre parcerias e investimentos em concorrentes que possam ser consideradas ameaças ao mercado do Sicredi

Objetivos 2019 foi...



Exploração de Inteligência Artificial em Inteligência Competitiva

[1] Filtros

Como podemos usar Inteligência Artificial para aprimorar os filtros de pesquisa de notícias para o Radar de Inteligência Competitiva?

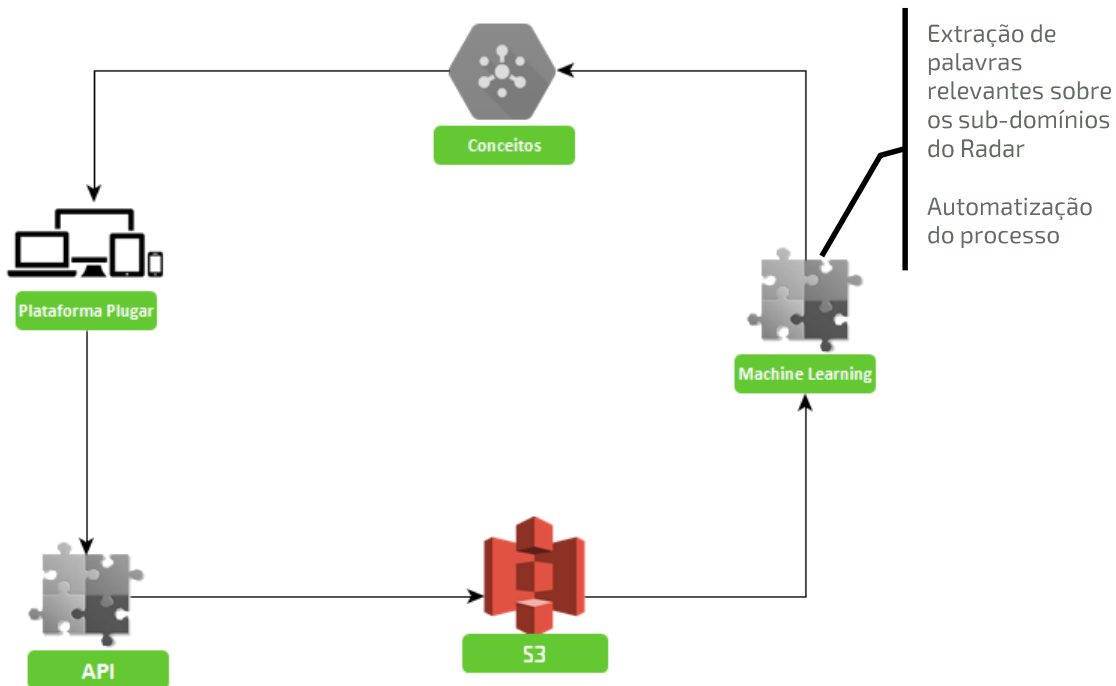
[2] Atualização

Como podemos proceder a atualização desses filtros de forma automática?

[3] Extração Relações

Através da IA realizar correlações entre parcerias e investimentos em concorrentes que possam ser consideradas ameaças ao mercado do Sicredi

Desenvolvimento



- **Uso de Machine Learning para gerar tópicos nos sub-domínios do Radar de Inteligência Competitiva;**
- **Possibilidade de atualização automática dos tópicos sem intervenção humana;**
- **Processamento automático possibilita o monitoramento de tendências;**

Machine Learning

LDA - Latent Dirichlet Allocation

- ✓ Técnica largamente usada para extração de tópicos
- ✓ Usa método estatístico – pode ser utilizado em qualquer língua
- ✓ Implementação usando Python – open source

Tópicos:

Tópico 1: plataforma + financeira + mobilidade + urbana
Tópico 2: abordagem + temas + mudanças + climáticas
Tópico 3: orientação + financeira + pessoa + física

Exemplos:

1

A Cabify e a BID Invest anunciam um financiamento de US\$ 70 milhões, concedida pela entidade financeira à plataforma de mobilidade urbana. (...)

2

O Itaú Unibanco revisitou a abordagem para tratar de temas relacionados às mudanças climáticas e lança seu posicionamento (...)

3

O Santander Brasil lançou uma plataforma gratuita e única no mercado para educação e orientação financeira para todos os clientes pessoa física (...)

Detalhes de implementação



- ✓ Código :
 - ✓ Desenvolvido em Python usando Anaconda/Spyder
 - ✓ Biblioteca de apoio: NLTK - <https://www.nltk.org/>
 - ✓ Importante: padronização e lematização das palavras (lemma: **banc** – banco,bancos,bancários,bancarizados,...)
- ✓ Parâmetros
 - ✓ Teste com vários parâmetros (tópicos e palavras)
 - ✓ O mais adequado se mostrou com a configuração de 20 tópicos e 3 palavras
- ✓ Quantidade Documentos: 6.000 notícias (últimos 3 meses)

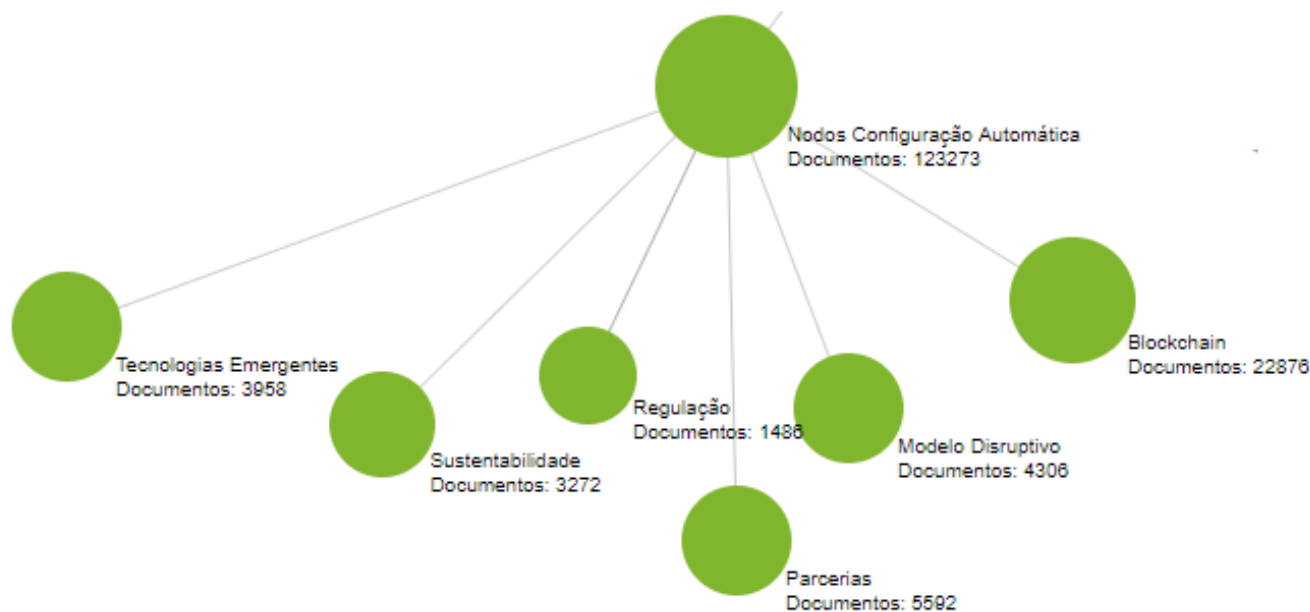


Resultados – Plataforma Plugar

Configuração dos nodos automáticos



PLUGAR
Data & Intelligence



Trabalhos futuros



- Testar num conjunto maior de notícias: 100.000
- Processar o experimento na Cloud AWS
- Aplicar a técnica para novos temas monitorados por outras áreas do Sicredi
- Identificar novos "temas/tendências" para monitoramento

Um exemplo de código pode ser encontrado em:



https://github.com/patriciangoncalves/LDA_ExtracaoTopicos



Agradecimentos

Iniciativa Inteligência Competitiva  **Sicredi**



Gustavo Aprile Porto Rossi
Especialista de Planejamento Estratégico



Guilherme Schaeffer



Carolina Bernardi



Matheus Quevedo



Patricia Nunes Gonçalves

Sicredi

QUARTA | 27/11

Trilha Inteligência Artificial

O mundo está mudando e a inteligência artificial e a grande responsável por essa revolução, que vai mudar praticamente tudo como conhecemos.

QUINTA | 28/11

Trilha Machine Learning

Machine Learning e Deep Learning além do buzzword

Muito se fala de Machine Learning, Deep Learning e seu poder supostamente revolucionário, mas poucas vezes se mostra o que há por trás desses



THE DEVELOPER'S CONFERENCE

QUINTA | 28/11

Trilha Arquitetura de Dados

Discutir e compartilhar experiências sobre arquitetura de dados RDBMS e NoSQL

Nos dias de hoje com a infinidade de tecnologias de



12:10
às
13:00

Experimento de Extração de Relações com algoritmo Conditional Random Field ▾

Patricia Nunes Gonçalves

Machine Learning Canvas: Da coleta de dados à geração de valor ▾

Alexandre Ray Silva



11:10
às
12:00

Usando LDA para seleção de tópicos para Recuperação de Informação aplicado a Inteligência de Mercado ▾

Patricia Nunes Gonçalves

OCR: o que é e como usar? ▾

Guilherme Malta



10:10
às
11:00

Um estudo sobre implementação de ILM (Information Lifecycle Management) no Sicredi ▾

Patricia Nunes Gonçalves

Kafka e Dataflow: uma história de amor ▾

Leonardo Alves Miguel

Perguntas?



THE
DEVELOPER'S
CONFERENCE



patricia_ngoncalves@sicredi.com.br



<https://dblp.org/pid/14/6534>



<https://www.linkedin.com/in/patricianunesgonc/>

