

Agrupamento de Textos e suas Aplicações em Inteligência Analítica

Contextualização e Motivação

Ricardo M. Marcacini
ricardo.marcacini@icmc.usp.br

Cursos de Extensão – Difusão de Conhecimento – Dezembro de 2021



Inteligência Analítica para Textos

■ Por qual motivo analisar dados textuais?

- Notícias
- Redes Sociais
- Artigos Científicos
- Boletins Financeiros
- Laudos, Livros, E-mails, Help-desk, etc...



Inteligência Analítica para Textos

■ Por qual motivo analisar dados textuais?

- Notícias
- Redes Sociais
- Artigos Científicos
- Boletins Financeiros
- Laudos, Livros, E-mails, Help-desk, etc...



Fonte: [Img1]

Textos representam 80% da informação
existente nas organizações!

Forma natural do ser humano em transferir conhecimento.

Inteligência Analítica para Textos

■ Por qual motivo analisar dados textuais?

- Notícias
- Redes Sociais
- Artigos Científicos
- Boletins Financeiros
- Laudos, Livros, E-mails, Help-desk, etc...



Fonte: [Img1]

Inteligência Analítica para Textos

Inteligência Analítica para Textos

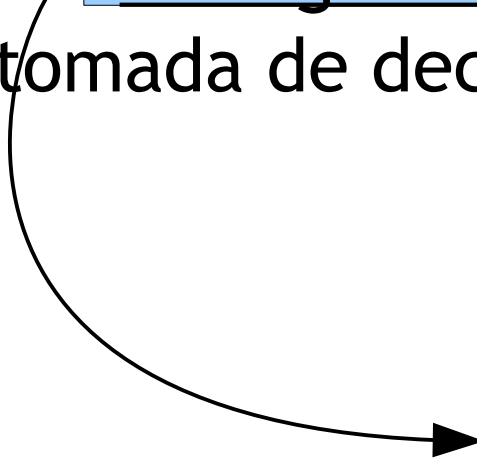
■ Definição

É um processo de extração de conhecimento útil a partir de grandes bases de textos usando com apoio de inteligência artificial para apoiar processos de tomada de decisão.

Inteligência Analítica para Textos

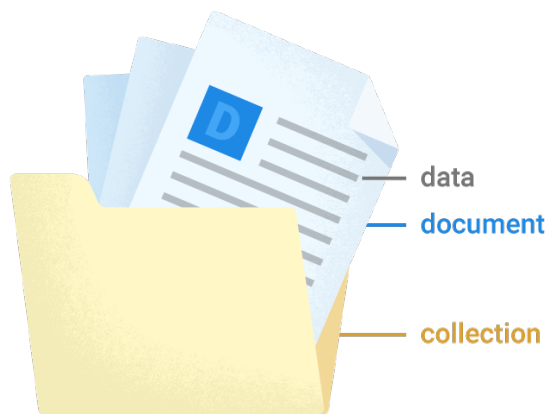
■ Definição

É um processo de extração de conhecimento útil a partir de grandes bases de textos usando com apoio de inteligência artificial para apoiar processos de tomada de decisão.

- 
- Aprendizado de Máquina
 - Mineração de Dados e Textos
 - Ciências de Dados e Big Data

Inteligência Analítica para Textos

- Quais os desafios?
 - Textos são dados NÃO ESTRUTURADOS!



Fonte: [Img2]

Textos

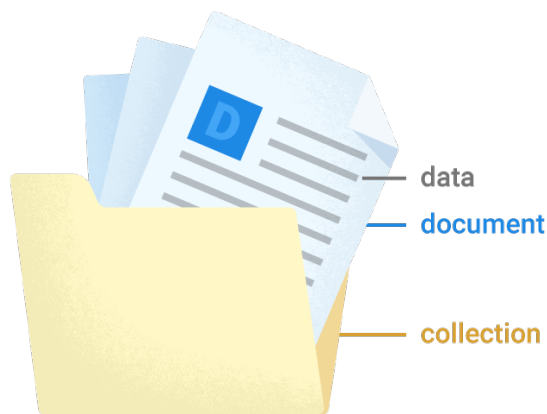


Tomada de Decisão

Inteligência Analítica para Textos

- Quais os desafios?
 - Textos são dados NÃO ESTRUTURADOS!

Como extrair conhecimento dos textos?



Fonte: [Img2]

Textos



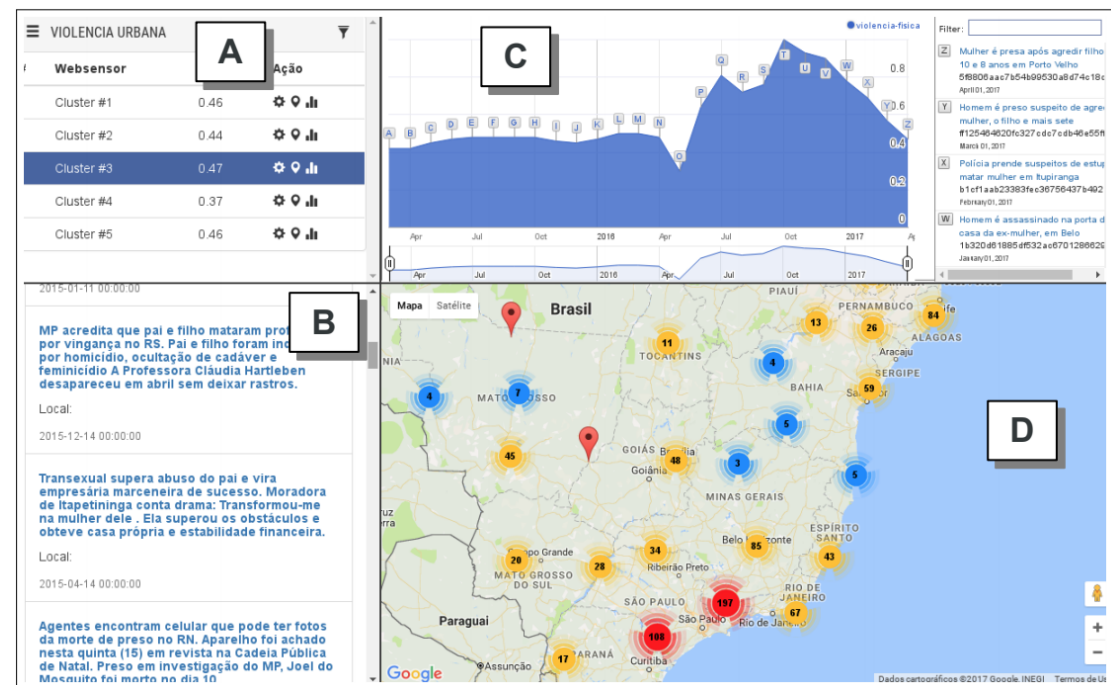
Tomada de Decisão

Inteligência Analítica para Textos

■ Exemplo e visão geral

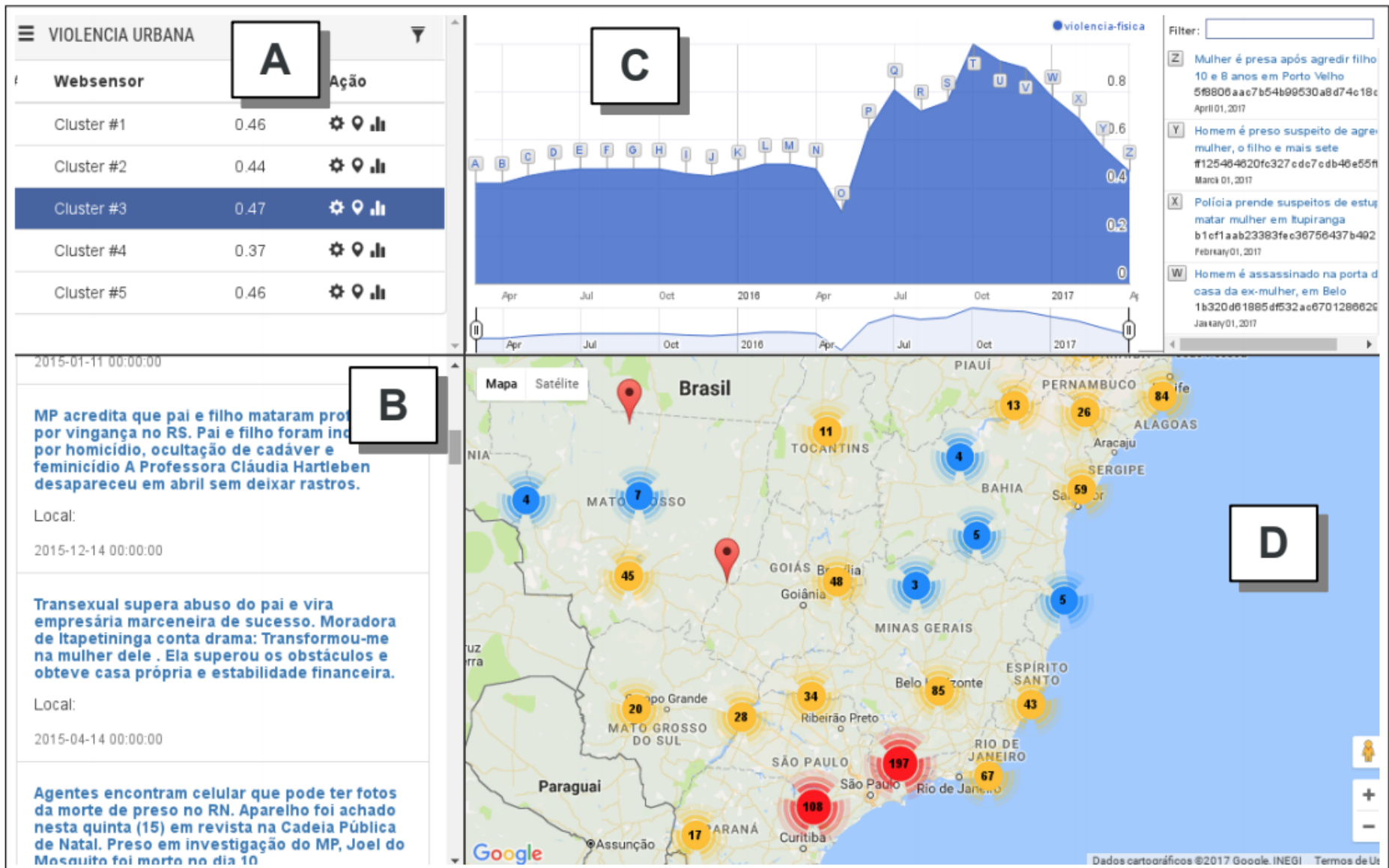
- A partir de uma amostra de textos de interesse, obter grupos de documentos relacionados
- Exemplo:

Projeto Websensors



MARCACINI, Ricardo M. et al. Websensors analytics: Learning to sense the real world using web news events. In: Anais Estendidos do XXIII Simpósio Brasileiro de Sistemas Multimídia e Web. SBC, 2017. p. 169-173.

Inteligência Analítica para Textos

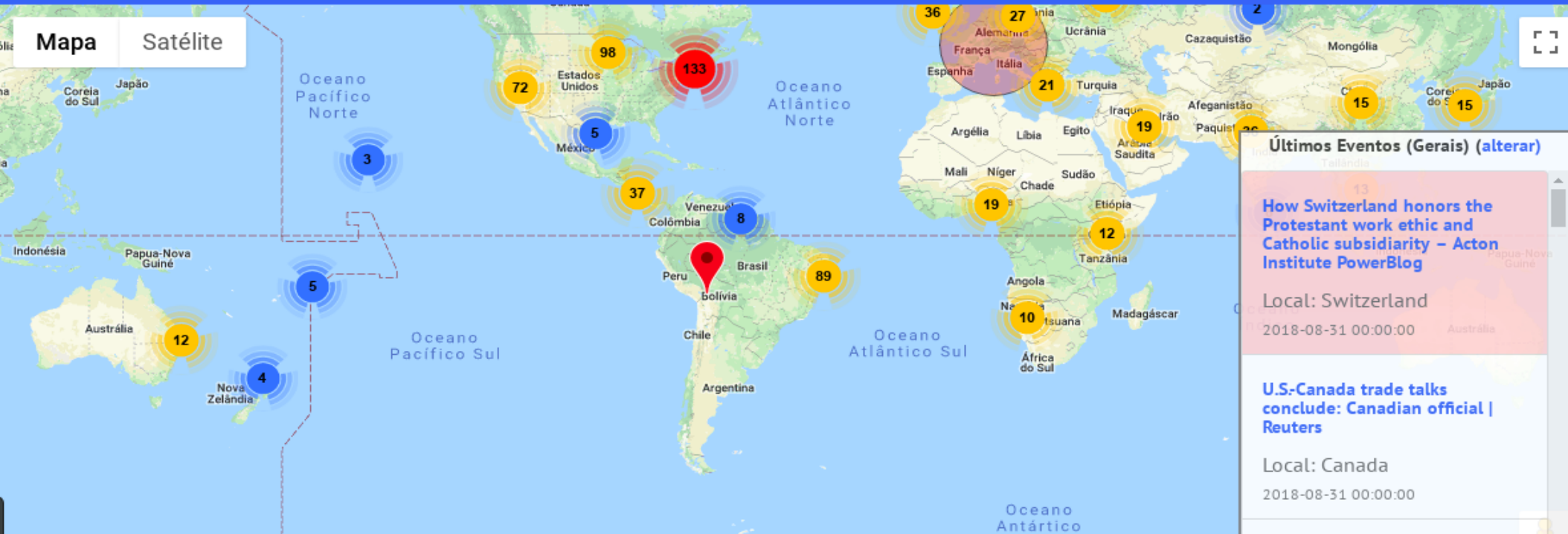


MARCACINI, Ricardo M. et al. Websensors analytics: Learning to sense the real world using web news events. In: Anais Estendidos do XXIII Simpósio Brasileiro de Sistemas Multimídia e Web. SBC, 2017. p. 169-173.

Exemplo (Websensors)

Mapear eventos que ocorrem na web (mundo virtual) para o nosso mundo real. Monitorar fenômenos, tendências e realizar tarefas descritivas e preditivas.

Coleta e monitoramento de eventos em tempo real.



Inteligência Analítica para Textos

■ Quais os desafios?

- Textos possuem várias componentes de informação que precisam ser extraídas

20/04/2017 (AgroNews)

O estado de Mato Grosso do Sul é o único que deve contabilizar aumento de produção de soja para o próximo ano, conforme relatório divulgado nesta quinta-feira pela Conab (Companhia Nacional de Abastecimento)

- O que aconteceu?
Componente WHAT

Inteligência Analítica para Textos

■ Quais os desafios?

- Textos possuem várias componentes de informação que precisam ser extraídas

20/04/2017 (AgroNews)

O estado de Mato Grosso do Sul é o único que deve contabilizar aumento de produção de soja para o próximo ano, conforme relatório divulgado nesta quinta-feira pela Conab (Companhia Nacional de Abastecimento)

- O que aconteceu?
Componente **WHAT**
- Onde aconteceu?
Componente **WHERE**

Inteligência Analítica para Textos

■ Quais os desafios?

- Textos possuem várias componentes de informação que precisam ser extraídas

20/04/2017 (AgroNews)

O estado de Mato Grosso do Sul é o único que deve contabilizar aumento de produção de soja para o próximo ano, conforme relatório divulgado nesta quinta-feira pela Conab (Companhia Nacional de Abastecimento)

- O que aconteceu?
Componente WHAT
- Onde aconteceu?
Componente WHERE
- Quando aconteceu?
Componente WHEN

Inteligência Analítica para Textos

■ Quais os desafios?

- Textos possuem várias componentes de informação que precisam ser extraídas

20/04/2017 (AgroNews)

O estado de Mato Grosso do Sul é o único que deve contabilizar aumento de produção de soja para o próximo ano, conforme relatório divulgado nesta quinta-feira pela Conab (Companhia Nacional de Abastecimento)

- O que aconteceu?
Componente WHAT
- Onde aconteceu?
Componente WHERE
- Quando aconteceu?
Componente WHEN
- Quem está envolvido?
Componente WHO

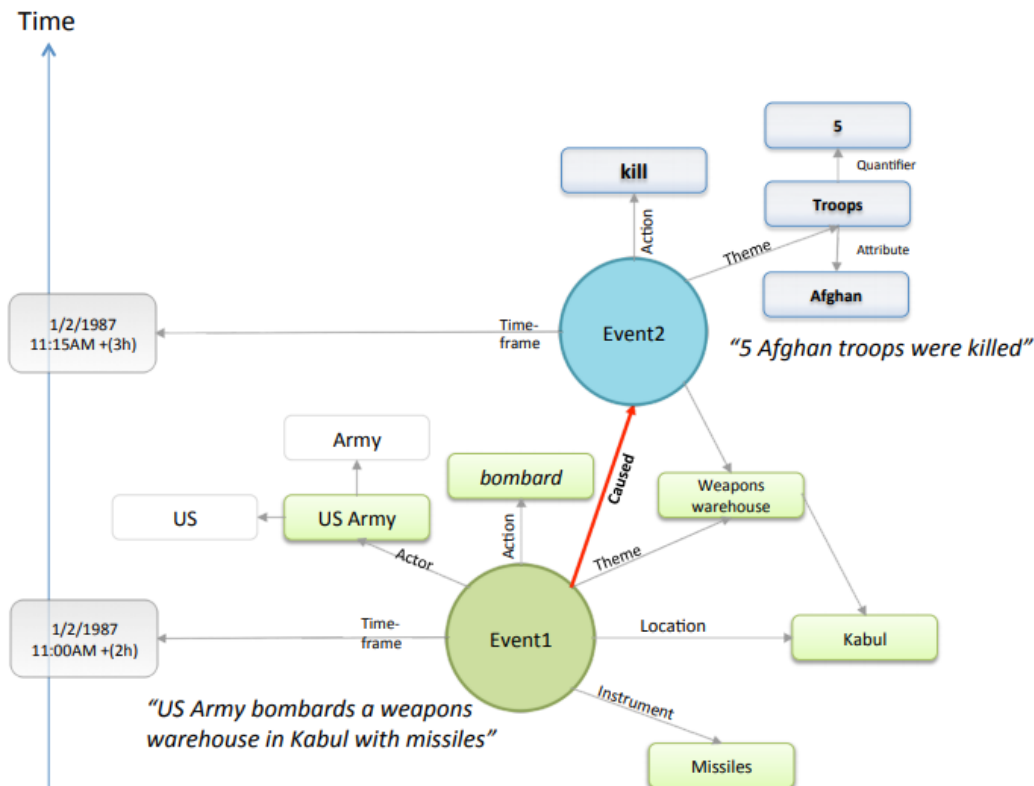
Inteligência Analítica para Textos

- Quais os desafios?
 - Identificar documentos relacionados por algum critério de similaridade

Inteligência Analítica para Textos

■ Quais os desafios?

- Identificar documentos relacionados por algum critério de similaridade

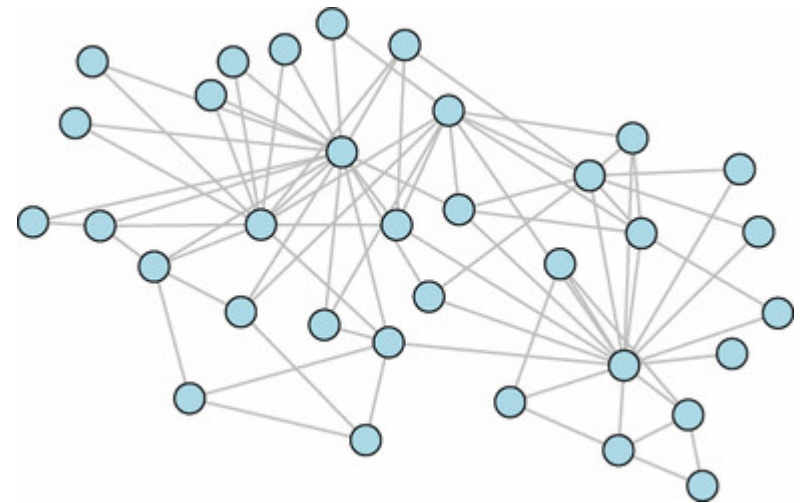
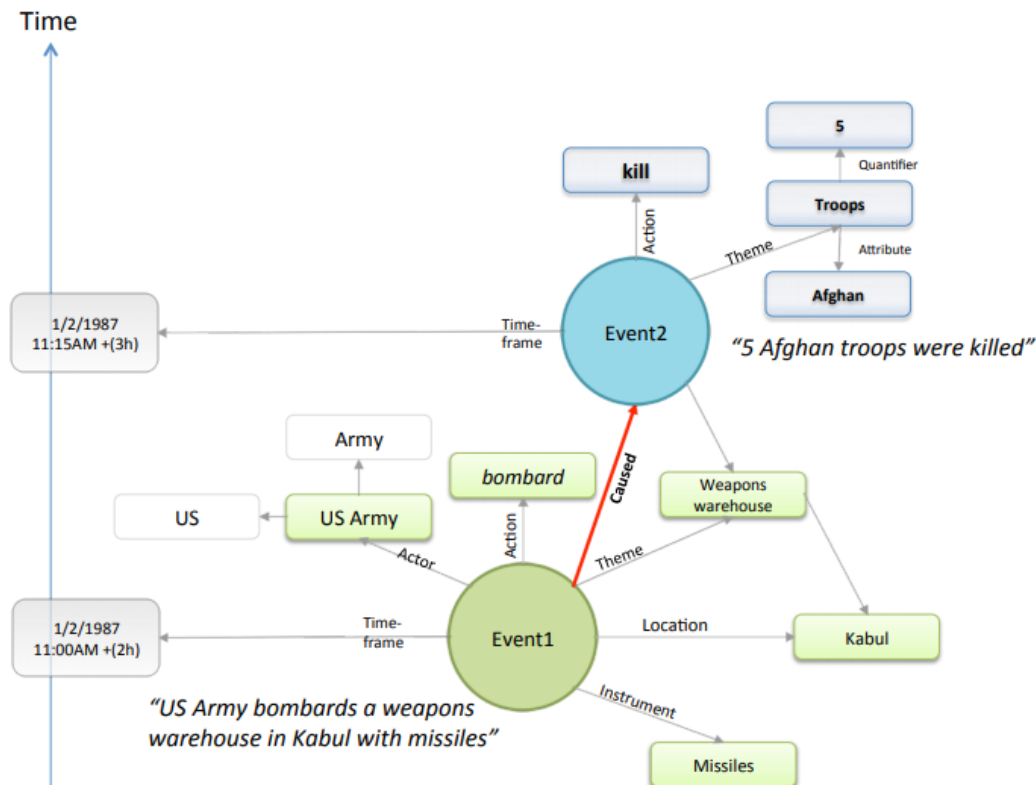


RADINSKY, Kira; HORVITZ, Eric. Mining the web to predict future events. In: Proceedings of the sixth ACM international conference on Web search and data mining. 2013. p. 255-264.

Inteligência Analítica para Textos

■ Quais os desafios?

- Identificar documentos relacionados por algum critério de similaridade

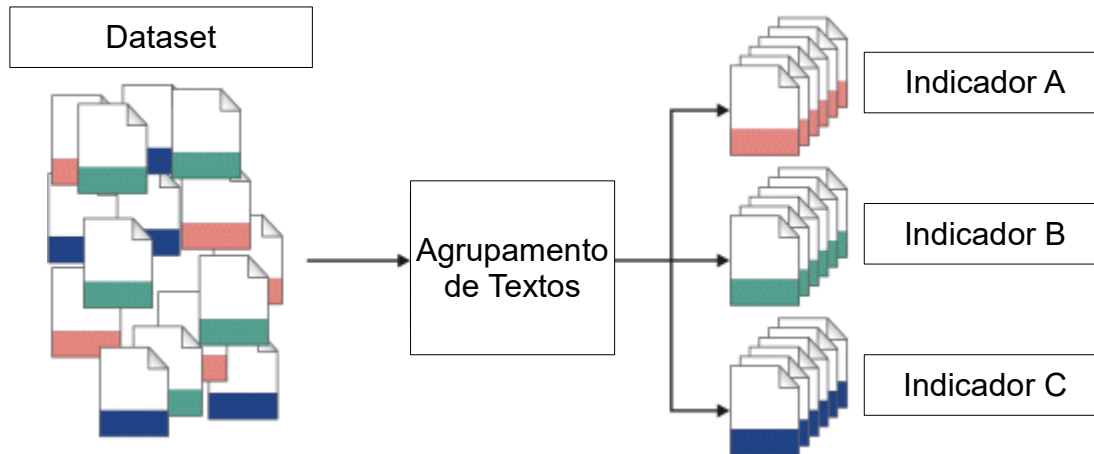


RADINSKY, Kira; HORVITZ, Eric. Mining the web to predict future events. In: Proceedings of the sixth ACM international conference on Web search and data mining. 2013. p. 255-264.

Inteligência Analítica para Textos

■ Quais os desafios?

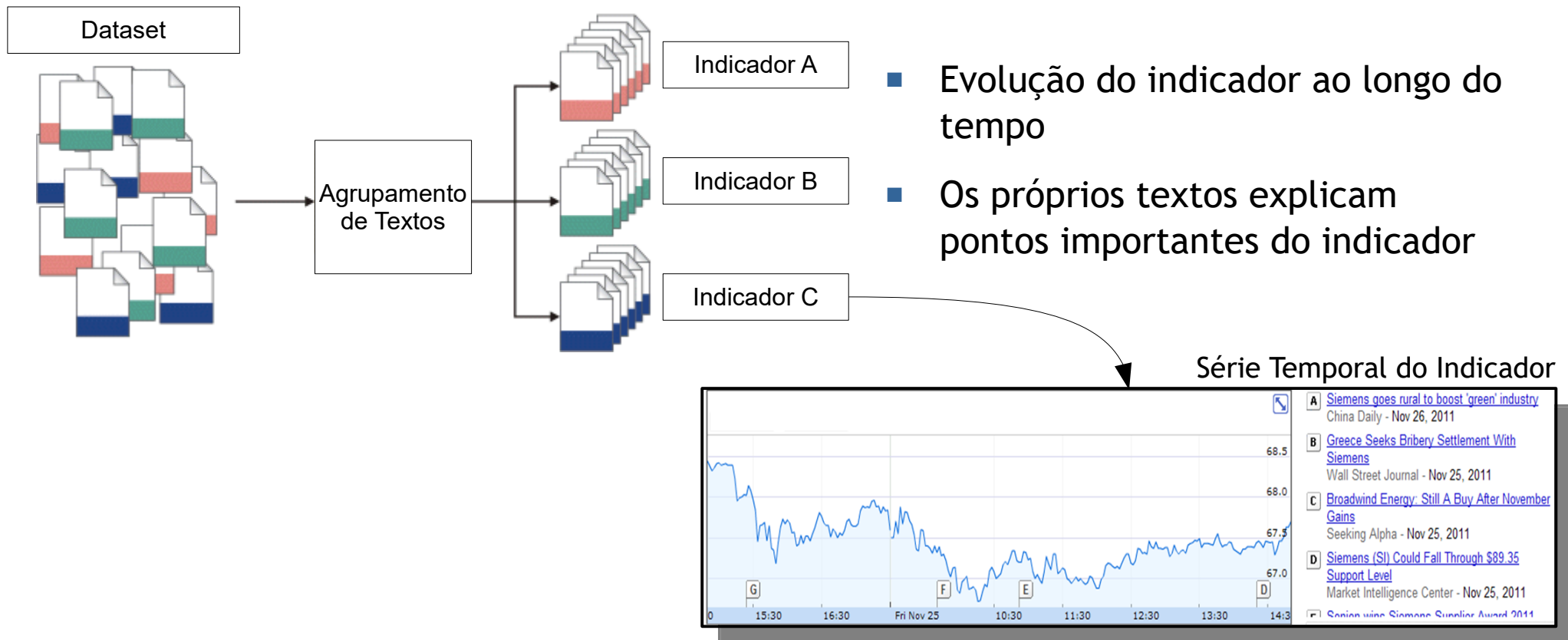
- Aprender indicadores (inteligentes) a partir de coleções textuais



Inteligência Analítica para Textos

■ Quais os desafios?

- Aprender indicadores (inteligentes) a partir de coleções textuais

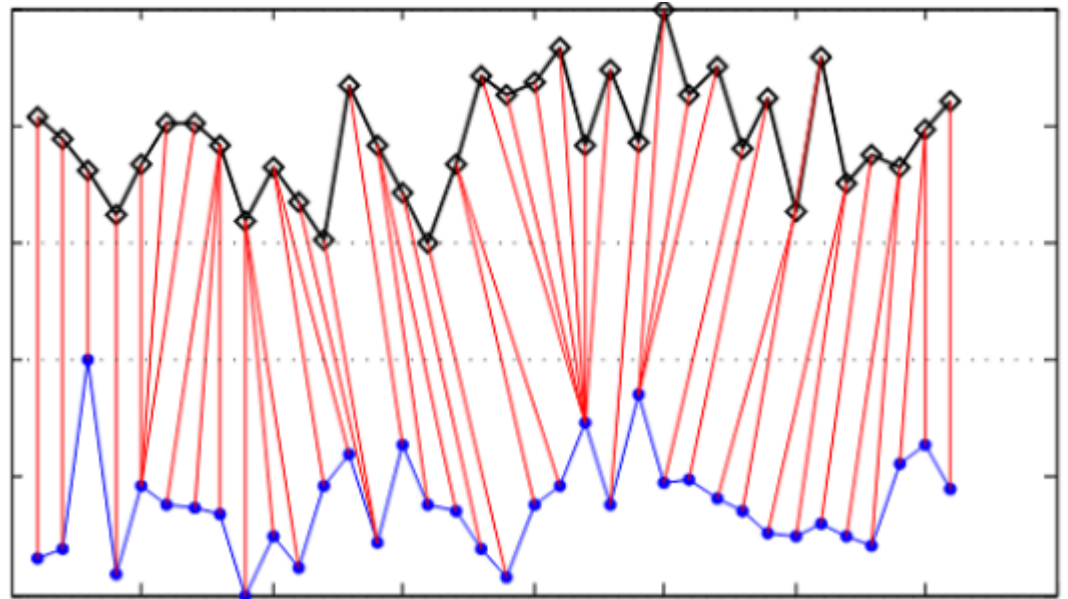
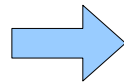


Inteligência Analítica para Textos

■ Quais os desafios?

- Correlacionar indicadores obtidos dos textos com “indicadores tradicionais”

- Indicador da empresa:
Ex: Taxa de Exportação de Produto

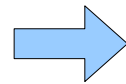


Inteligência Analítica para Textos

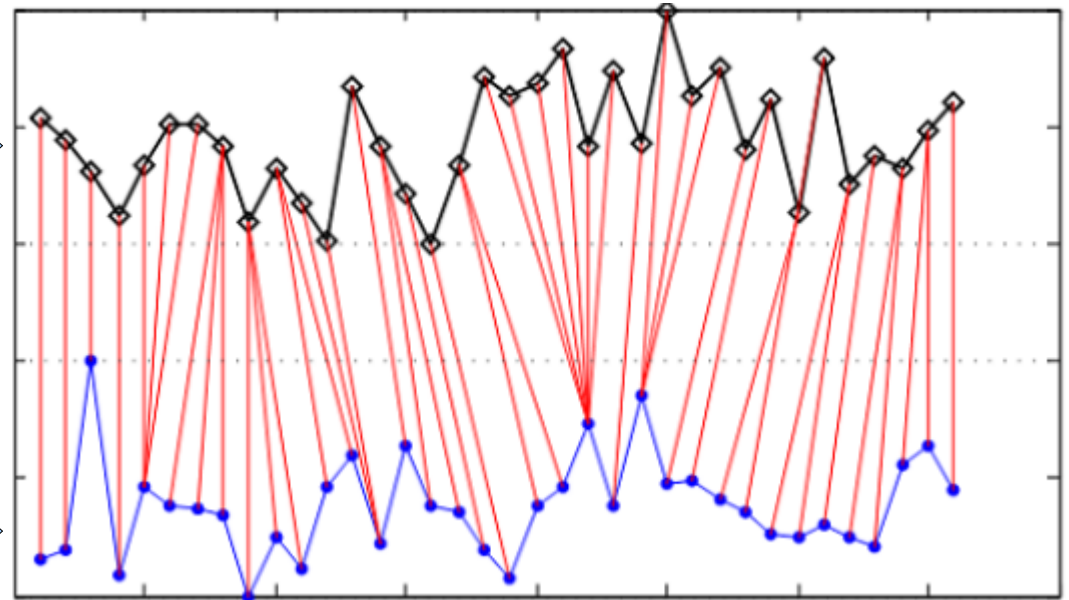
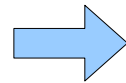
■ Quais os desafios?

- Correlacionar indicadores obtidos dos textos com “indicadores tradicionais”

- Indicador “aprendido”:
Ex: Eventos sobre Investimento em Tecnologia



- Indicador da empresa:
Ex: Taxa de Exportação de Produto

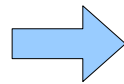


Inteligência Analítica para Textos

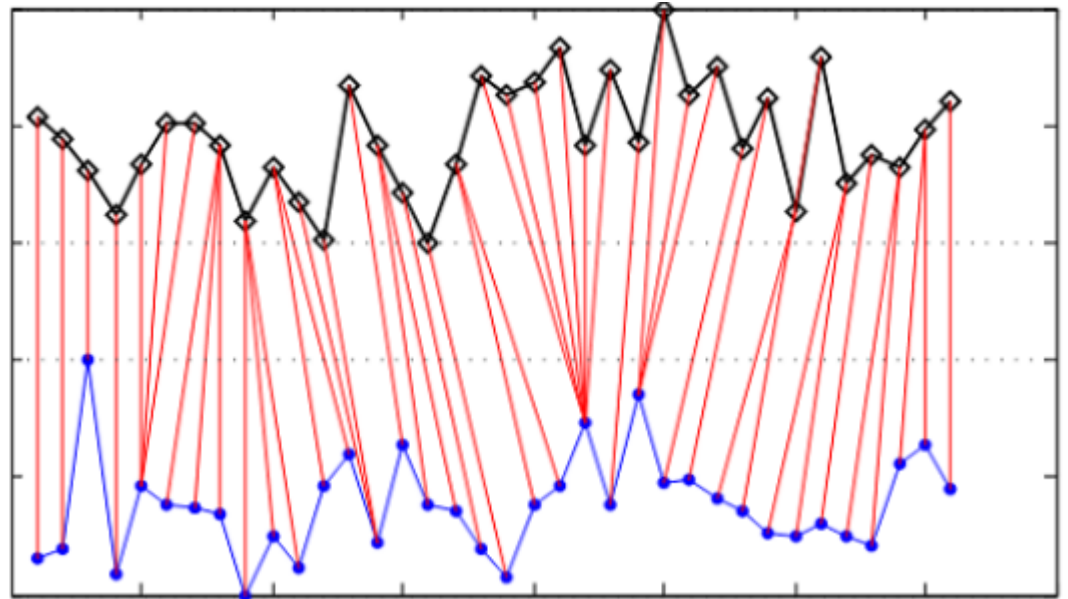
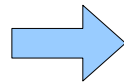
■ Quais os desafios?

- Correlacionar indicadores obtidos dos textos com “indicadores tradicionais”

- Indicador “aprendido”:
Ex: Eventos sobre Investimento em Tecnologia



- Indicador da empresa:
Ex: Taxa de Exportação de Produto

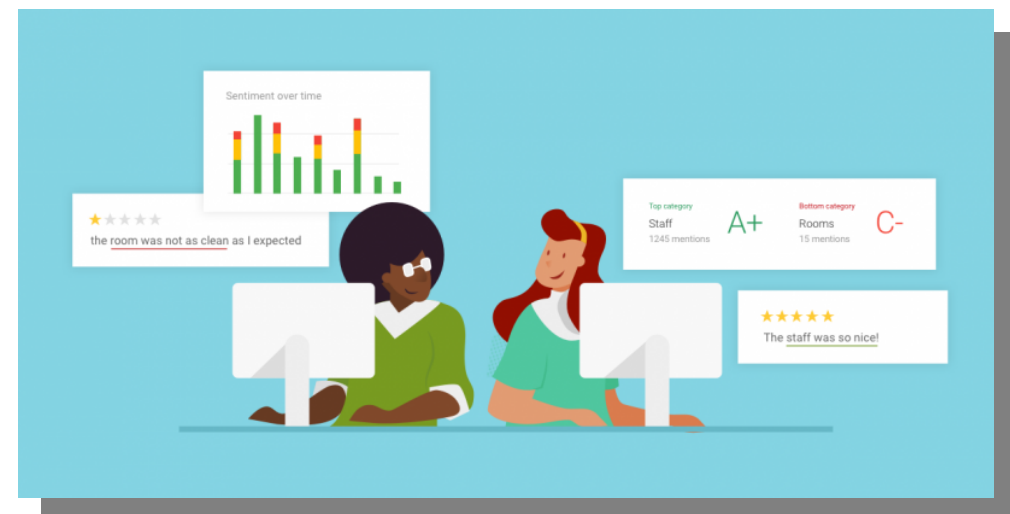


Apoio para tarefas descritivas (entender o comportamento passado) e tarefas preditivas (estimar comportamento futuro)

Inteligência Analítica para Textos

■ Análise em Redes Sociais

- Organizar posts e comentários em tópicos
 - WHO? → Consumidor, Empresa, etc.
 - WHAT? → É a opinião a respeito de algo/alguém.
 - WHEN? → Quando a opinião foi emitida.
 - WHERE? → Local do usuário, empresa, etc.



Birdeye.com, "The benefits of customer sentiment analysis tools." <https://s3.amazonaws.com/blog4.0/wp-content/uploads/2018/04/customer-sentiment-analysis-tools-810x405.png>

Inteligência Analítica para Textos

■ Marketing

- Tradicional: organizar clientes (segmentar) com base em suas preferências textuais.
- Identificar “oportunidades” para ações de marketing
 - Provável desistência?
 - Reclamando do concorrente?

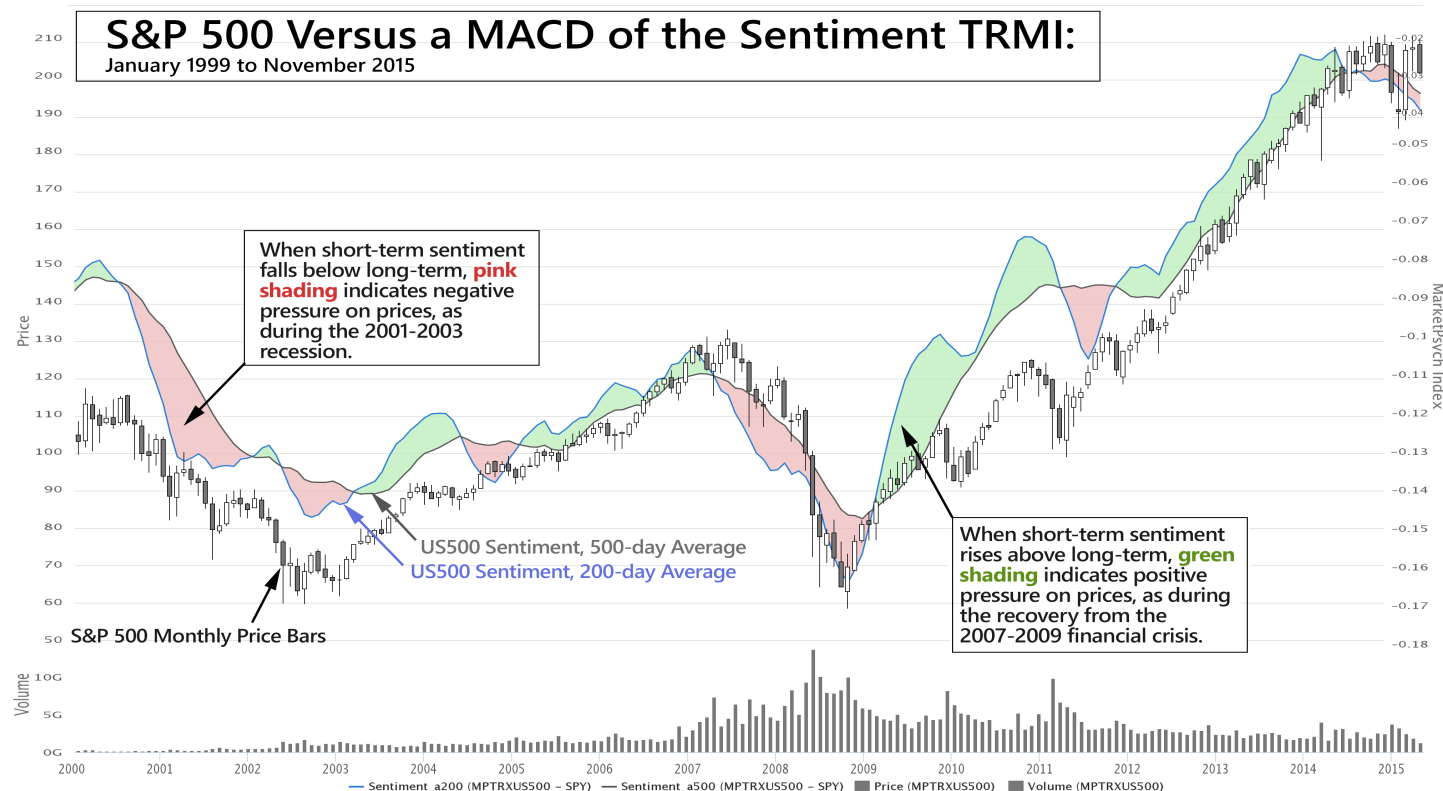


Marketing Insights, “Everything you need to drive growth through digital marketing.”
<https://www.smartinsights.com/wp-content/uploads/2012/05/social-network-communities-image.jpg>

Inteligência Analítica para Textos

■ Mercado Financeiro

- Organizar notícias e relatórios que podem afetar (positivamente ou negativamente) um ativo financeiro

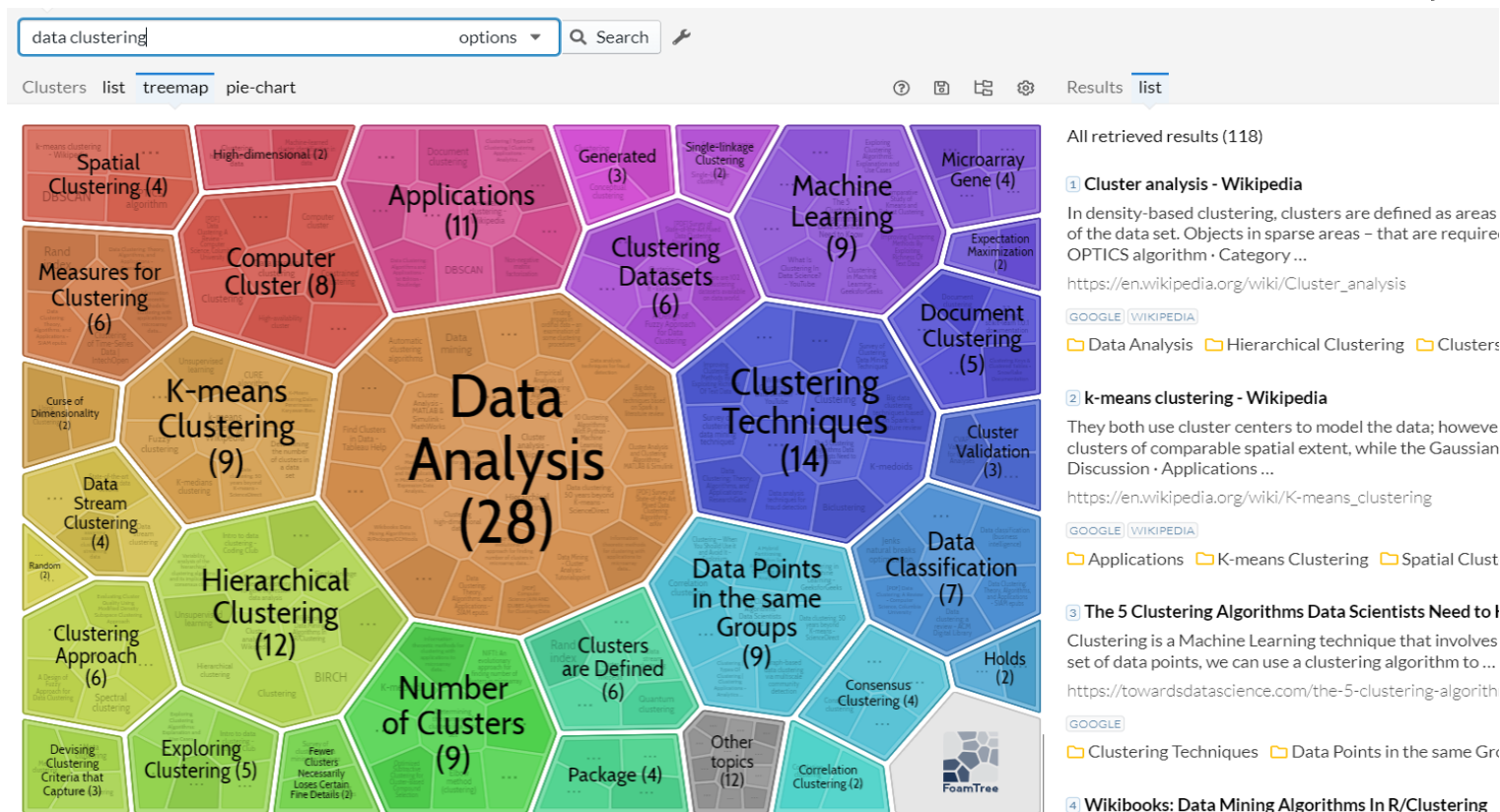


Marketpsych.com, "MarketPsych Report: How Fund Managers Trade on Sentiment."
http://marketpsych-website.s3.amazonaws.com/images/S%26P500_MACD_1998-2015.png

Inteligência Analítica para Textos

■ Análise Exploratória

- Organizar uma base textual em grupos e subgrupos facilita análise em diferentes níveis de abstração



Exemplo: <https://search.carrot2.org/>