



## ZYTE PAPER

### Data use cases

# Fueling business intelligence with web scraped news data

Three use cases demonstrating the power of web scraped news and article data in investing, sentiment analysis and competitor monitoring.



# Introduction

**Scraping data from news and online articles is a burgeoning practice enabling companies to power a wide array of business processes, giving them a competitive edge in the market.**

For traditional businesses, technology companies, hedge funds, political campaigns, and non-profits, scraping web data from online articles and news sources is akin to having a finger on the pulse of their reputation, competition and industry.

Public relations officers and members of executive management alike don't have time to seek out and pursue every granular reference to their or their competition online, so by automating these processes they can receive the most important, mission-critical updates faster than their competitors.

As a web scraping pioneer, we work with thousands of businesses and scrape tens of billions of data sources a month, giving us a unique perspective of the news and article monitoring industry and the fascinating ways in which they use web scraped data to power meaningful and effective business solutions.

In this guide we'll give you a sneak peek into four of the most common and impactful use cases businesses use to produce actionable insights and keep a watchful eye over the health and competitiveness of their industry.

Since online articles and news data is continually produced at an exponential rate, the usefulness for this data is expected to increase.

We hope you'll find an exciting new use for web scraped news and article data in the following use cases:



## Investment decision making



## Online public sentiment analysis



## Competitor monitoring



## Political campaigns

# Investment decision making

**At Zyte, we've seen the rise of alternative data first-hand, with financial firms and hedge funds increasingly requesting and consuming non-traditional forms of data, and integrating it into their decision making processes for the treasure trove of insights it produces.**

This data is fueling a true disruption of the investment research landscape and changing the way firms seek out and generate actionable insights to the extent in which RoD - return on data - has become a common turn of phrase.

Extracting insightful data from online news sources and industry articles enables account managers to closely monitor their investments and ensure their decisions are being made from the most informed position possible. By coupling this data with natural language processing and AI, account managers are able to turn industry commentary, company press releases and news into leading indicators of a companies performance.

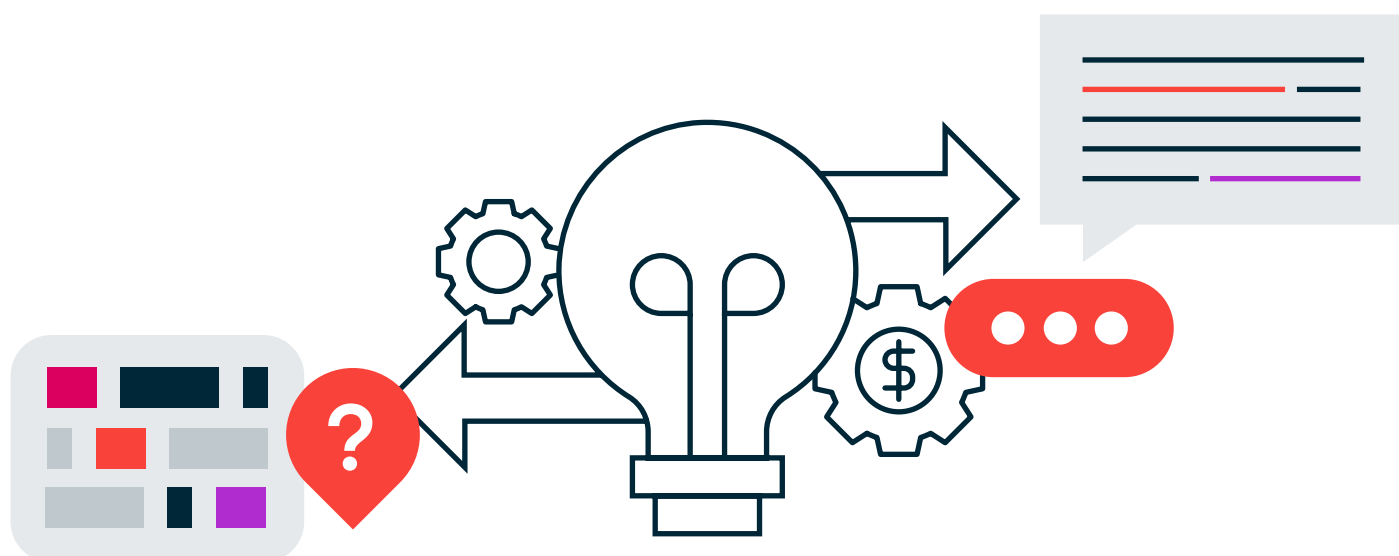
Furthermore, research in the financial industry has shown that news articles and social media opinions can directly influence the stock market. Both the informational and subjective aspects of online articles can impact volatility, stock price, trading volume, and market activity.

The efficient market hypothesis suggests that investors who are able to gain access to online public sentiment are able to make better trades and better maximize their returns.

## Powering profitable selection

Data that tracks sentiment signals is efficacious in the stock selection process, and some studies have found sentiment as effective or more effective than price momentum in real time.

Traditional reliance on sell-side analysis may be limited compared to the potential that lies in large scale automated sentiment extraction from news stories.





### Timing entry

- Announcement data provides an advantage over reported CapEx.



### Expected returns across a portfolio

- Machine learning technologies have utility in developing alternative risk premia strategies. Many investors want to move away from traditionally static risk allocation towards a strategy that includes valuation-based tilts designed to outperform the market.
- Web scraped data is key in developing more nuanced risk allocation structures. Whether the goal is a tilt toward higher volatility asset classes or risk aversion, automated news data extraction and analysis aid in maintaining the strategy.



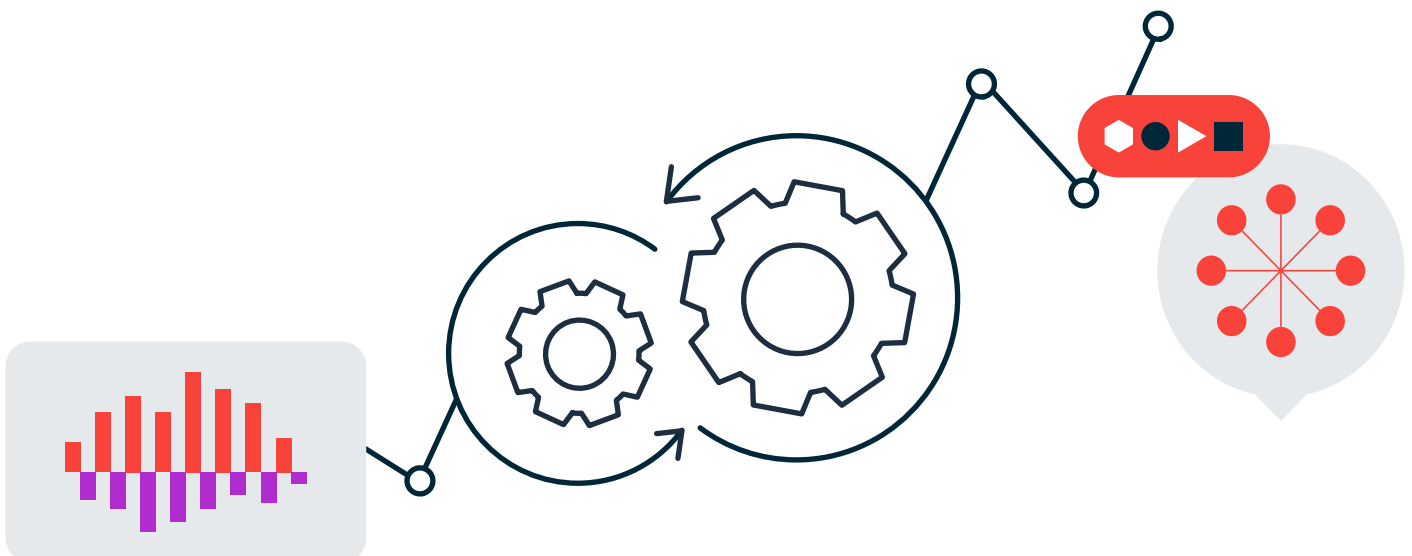
### Predictive of future firm fundamentals

- Sentiment analysis from web scraped news data has been shown to be predictive of net income, cash flows, and dividends.
- The emotion lexicon allows machine learning technologies to categorize large amounts of data as positive or negative and track changes in the sentiment a company generates in aggregate.



### DJIA correlation with sentiment

- Dow Jones Industrial Average values are correlated with “calm” and “happy” sentiment states, so there is real predictive power in this data for buy and sell decisions.



# Measuring online public sentiment

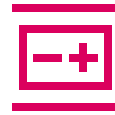
**A business intelligence (BI) strategy is incomplete without an understanding of public sentiment. The way a company, industry, or product is portrayed in news articles and reviews directly impacts a company's bottomline. Sentiment analysis is the process of understanding public opinion from written data.**

By scraping sentiment data, business intelligence teams can both capitalize on positive publicity and work to mitigate negative sentiments. A comprehensive BI strategy acknowledges the multipronged nature of online sentiment by scraping data from many sources and analyzing many of the factors that comprise sentiment. There is simply too much data for people to analyze without utilizing automated scraping technologies.

Automation helps to eliminate the biased aspect of sentiment analysis and allows businesses to reference a much larger data set. Machine learning text analysis and online data extraction are two of the tools enabled by scraping technologies.

## Powering AI-driven analytics

Machine learning capabilities allows businesses to transform the vast amount of textual data on review websites, forums, and articles into actionable insights



### Polarity analysis

- Sentiment analysis tools can give a high level view of product or company perception by tracking polarity and whether overall reviews are “good” or “bad” to put it simply.
- Real time polarity analysis can help businesses project how a product is trending.



### Aspect based Sentiment analysis

- Product feedback can be categorized on text-based tags associated with a specific topic or aspect of the company/product.
- This machine learning technique can break down text into aspects and yield insight into how customers feel about specific attributes of a product. The analysis can show us the patterns of specific sentiment frequency as they relate to a specific product component. Aggregate analysis may be positive, but a certain feature tied to negative sentiment is an opportunity for product and company improvement.



## Benefits of Automated Sentiment Analysis

- **Scalability:**

The sheer amount of product reviews, articles, and online opinion data renders it impossible to manually process the available data. AI and machine learning automation can get us a step closer to analyzing a vast amount of the data.

- **Consistent criteria:**

Applying the same tags and sentiment scores to various sources allows these automated processes to assess opinion with the same lens, analyzing a vast amount of the data.

- **Real time:**

Taking action swiftly can make the difference between a small blip and a widespread PR crisis.

- **Deeper insight:**

Human error may cause someone reading through reviews to categorize something as positive or negative without delving into each sentence. Automated processes can give business more in-depth insight into what is being said.



# Competitor monitoring

**The traditional non-data driven competitor monitoring processes employed by many businesses puts them at real risk for disruption. Creating blindspots and pain points for competitors to exploit.**

By integrating web scraped data into a systematic competitor monitoring process, companies gain a powerful advantage in their market and act quickly on competitor insights to maximize revenue, market share and exploit growth opportunities.

## **Ahead of the pack**

Thanks to the acceleration and proliferation of social media and digital news, articles can travel extraordinarily quickly and, as a result, are exponentially valuable data sources for competitive intelligence. By web scraping news and article data related to their competition, companies can use the aggregated intelligence to forecast product launches, analyze trends, and calculate opportunities for disruption.

Scraping competitor news, advertisements, sponsored posts, and corporate blog posts can reveal a great deal of insight into not only the segments they serve, their internal objectives and strategies, whether executive or promotional, but also insights into value migrations.

Understanding key value migrations within the increasingly transitory nature of today's business giants unlocks huge potential in aligning company strategy with disruptive business practices.

Corporations love to talk about themselves, and while many become tight-lipped around quarterlies, much can be revealed by analyzing aggregated news data. For example, those monitoring Microsoft over the past few years may have noticed events precipitating their 2018 reorganization, which represents a big value migration towards cloud based services.

## **Solutions for every strategy**

Web scraped news and article data empowers competitive intelligence practices by complementing core competencies and revealing a great deal of insight into competitor activity.

Most competitor monitoring strategies assign particular weights to different variables of a competitor's performance, creating datasets which can be greatly buttressed by a consistent stream of web scraped data. This data can inform analysis and continuously notify users with updates on numerous aspects of their competition, including:



### Segmentation strategy

- Corporate news articles are often written with an audience firmly in mind and conspicuous in writing - aggregating and analyzing textual data, especially with AI, may yield insights into changing market segmentation strategies.



### Distribution strategy

- Corporate media can betray changes in distribution strategy, whether it be an intensification of efforts in a particular region or a new exclusive distribution deal.



### Product line extensions and contractions

- By scraping news and articles related to a competitor's product line, their product line strategies become more clearly



### Audience interests

- The aggregation and weighting of textual data regarding the competitor itself may yield insights into what their audience truly demands.





# Political campaigns

**Constituent relations are critical to the health and success of any organization dealing with the public, and new solutions driven by web scraped news and article data empowers non-profits, NGOs, elected and aspiring public representatives to most effectively and meaningfully represent and engage their base.**

The diversity with which we've seen this data used in campaigning is remarkable, with quantity consumed and potential uses proliferating each day. The demand for this data is sky rocketing thanks to its sheer effectiveness, which, without exaggeration, is vital to the performance of any agent or organization involved with a constituent base.

## Everywhere all the time

Government relations officers, nonprofits, and politicians take full advantage of the tremendous utility of web scraped news data, using it to be first-to-react to issues important to constituents, respond quickly to breaking news cycles, and stay fully tuned into the public's reception of pending policy changes.

Internationally businesses rely on the timely, reliable nature of scraping news and article data to keep track of the myriad moving pieces of legislation related to their industry.

The drinks industry, for example, heavily utilizes web scraping to ensure changing laws in their countries of distribution don't harm revenue.



## Powering democratic engagement in the modern era

The rapid development and deployment of technology across the world's free democracies has posed unique challenges to the way we govern and engage with our representatives. Fortunately, utilizing web scraped news and article data can be used in the true pursuit of a more vibrant, robust democracy, driving true innovation in the following ways and many more:



## Political

- Enables more accurate and timely representation of constituent issues
- Ensures stories and developments affecting voters are consistently tracked by constituent relation manager
- Empowers federal-level political actors and institutions to properly execute oversight obligations



## Governmental

- Fuels cutting-edge processes to detect and combat propaganda from foreign state apparatuses
- Assists officials in tracing misinformation campaigns to their roots and litigating against fraudulent 'fake news' outlets



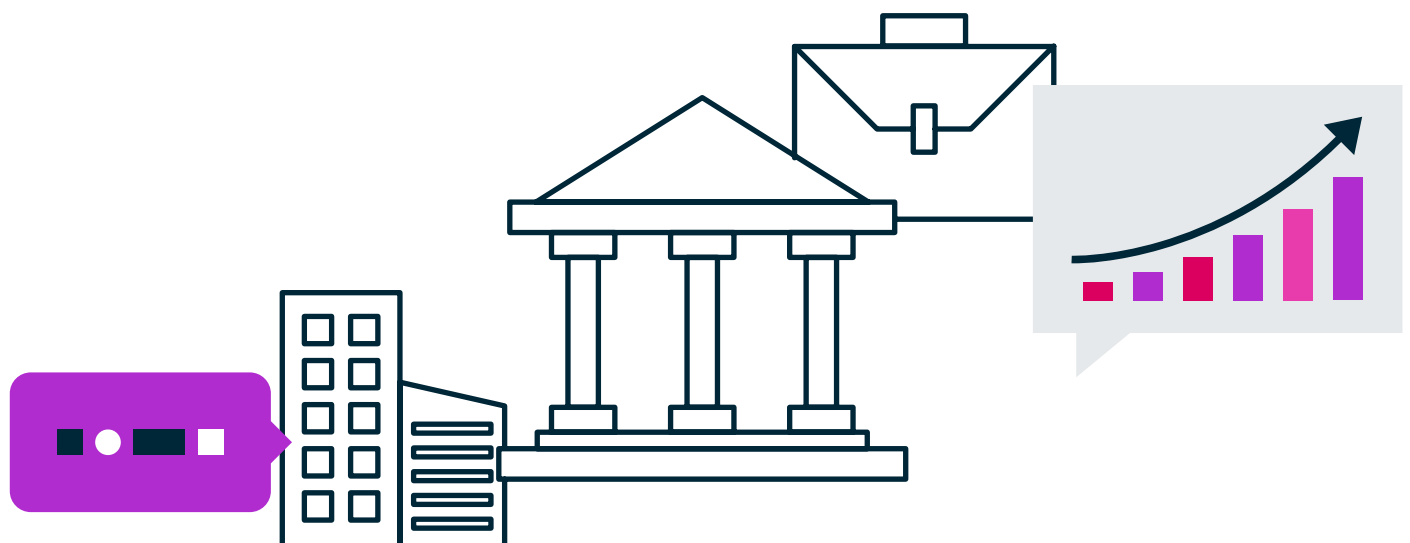
## Nonprofit

- Issue tracking enables nonprofits to establish and maintain topical authority and relevance
- Empowers legislative activism by informing lobbyists and public relations officers to demonstrate tangibility and relevance of topic to target committees



## Corporate

- Corporate affairs executives can monitor public or journalistic perspectives of legislative influence campaigns and scale operations accordingly



# Conclusion

**We hope some of these exciting use cases for news and article data have inspired you to implement such data into your own workflows and to innovate in your industry.**

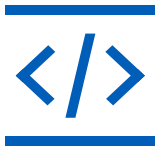
Investors, government officials and businesses alike can utilize news and article data to realize their fullest potential and endow their operations with the most effective, world-class solutions currently available.

As the internet continues to grow, the amount of data it generates grows with it, opening new opportunities for all types of organizations to improve their processes and make more informed decisions.

Therefore, we firmly believe that now is the best time to act, and that by incorporating such data streams into your organizational processes you can ensure your organization is disruption-proofed and fully prepared for the world of tomorrow.

Zyte is proud to offer the most robust web scraping solutions on the market, so if these use cases inspired you or you're curious about how web scraped news and article data can benefit your operation, then schedule a free consultation with our solution architecture team today.





At Zyte we turn websites into data with industry leading technology and services.

Our solutions include:

- **Data Extraction Service**

Let our web scraping experts build and manage the bespoke data extraction solution for your business needs.

- **Smart Proxy Manager (formerly Crawlera)**

Forget about proxy lists. We manage hundreds of thousands of proxies, so you don't have to.

- **Automatic Extraction powered by AI**

Instantly access accurate web data through our user-friendly interface or various Extraction APIs and save time getting the data you need.

- **Data extraction platform**

Access developer tools, data extraction APIs and documentation, built and maintained by our world-leading team of over 100 extraction experts.



It's yours. The web data you need.

Access clean, valuable data with web scraping services that drive your business forward.

[Talk to us](#)