



Investor Insight

**Exploring Companies' Annual SEC Reports through
Knowledge Graphs**

Akaash Dash, Aditya Natham, Sapan Patel, Max Zhao

SEC 10-K Filing

- Annual reports that publicly traded companies are required to file with the SEC
- Comprehensive - contains relevant background, financial data, risks, and opportunities
- Publicly available through the SEC

DIS > 10-K > 2018 > 2018 FY Annual report Filed: 21 Nov 18

**UNITED STATES
SECURITIES AND EXCHANGE COMMISSION**
Washington, D.C. 20549
FORM 10-K
ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE
SECURITIES EXCHANGE ACT OF 1934

For the Fiscal Year Ended September 29, 2018 Commission File Number 1-11605

The WALT Disney Company

Incorporated in Delaware
500 South Buena Vista Street, Burbank, California 91521
(818) 560-1000 I.R.S. Employer Identification No.
95-4545390

Securities Registered Pursuant to Section 12(b) of the Act:

Title of Each Class Name of Each Exchange
on Which Registered
Common Stock, \$.01 par value New York Stock Exchange

Securities Registered Pursuant to Section 12(g) of the Act: None.

Indicate by check mark if the registrant is a well-known seasoned issuer, as defined in Rule 405 of the Securities Act. Yes ☒ No ☐

Indicate by check mark if the registrant is not required to file reports pursuant to Section 13 or Section 15(d) of the Act. Yes ☐ No ☒

Indicate by check mark whether the registrant (1) has filed all reports required to be filed by Section 13 or 15(d) of the Securities Exchange Act of 1934 during the preceding 12 months and (2) has been subject to such filing requirements for the past 90 days. Yes ☒ No ☐

Indicate by check mark whether the registrant has submitted electronically and posted on its corporate Web site, if any, every Interactive Data File required to be submitted and posted pursuant to Rule 405 of Regulation S-T during the preceding 12 months (or for such shorter period that the registrant was required to submit and post such files).
Yes ☒ No ☐

Indicate by check mark if disclosure of delinquent filers pursuant to Rule 405 of Regulation S-K is not contained herein, and will not be contained, to the best of registrant's knowledge, in definitive proxy or information statements incorporated by reference in Part III of this Form 10-K or any amendment to this Form 10-K. ☐

Indicate by check mark whether the registrant is a large accelerated filer, an accelerated filer, a non-accelerated filer, a smaller reporting company or an emerging growth company. See the definitions of "large accelerated filer", "accelerated filer", "smaller reporting company", and "emerging growth company" in Rule 12b-2 of the Exchange Act.

Large accelerated filer	<input checked="" type="checkbox"/>	Accelerated filer	<input type="checkbox"/>
Non-accelerated filer	<input type="checkbox"/>	Smaller reporting company	<input type="checkbox"/>

Read sec 10K reports faster. CapEdge. (n.d.). Retrieved April 14, 2023, from <https://capedge.com/features/10k-annual-report-analysis>

The Problem

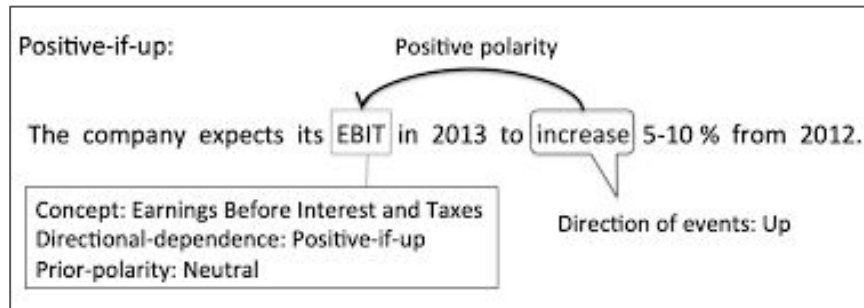
- 10-Ks are much too complicated and time consuming for a retail investor to read
- 'Information Gap' between 10-Ks and key content
 - Investor inequality
- Over time, companies started padding 10-Ks to appease institutional investors¹
 - Increases difficulty for retail investors



MSFT - # of words in 10-K filings vs year

Past Approaches

- Sentiment analysis
 - Doesn't allow investors to make their own decisions
 - Oversaturated, market overreacts to sentiment ²
 - Companies are aware of this and adjust their wording to ensure sentiment analysis indicates positive ³





Our Approach - SEC 10k Webpage Overview

- User enters ticker and year of a company
- Website retrieves 10-K, analyzes document, and generates word cloud and knowledge graph
- Quick method for investors to gain insight into company's operations
- Enables investors to make their own decisions

Search for a company

Company Ticker

e.g. AAPL

Document Year

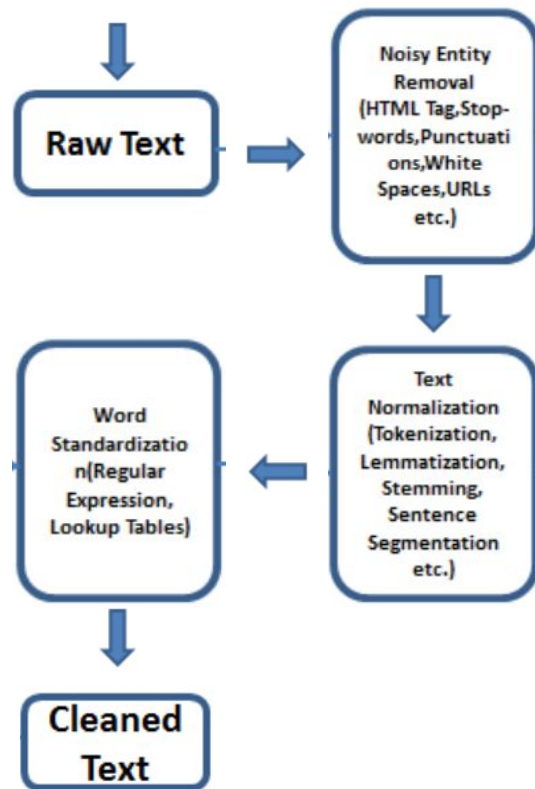
e.g. 2021

Go!

Web Page UI that allows user to enter Company's Ticker and year

Data Processing Pipeline

- Retrieve 10-K
 - Access SEC's CIK database to get CIK from ticker
 - CIK used on EDGAR to find company and 10-K for specified year
- Clean document
 - Remove numeric data, XML tags, whitespace, special characters
- Extract specific sections



Illustrating Word Cloud

- Create word cloud based on frequency
- Highlights focuses of company
 - Actions, nouns, etc.
- Simplistic snapshot without entity and relationships

Word Cloud of JPM Annual Report 2022



Word Cloud generated from JPM's 2022 10-k Filings



Flair

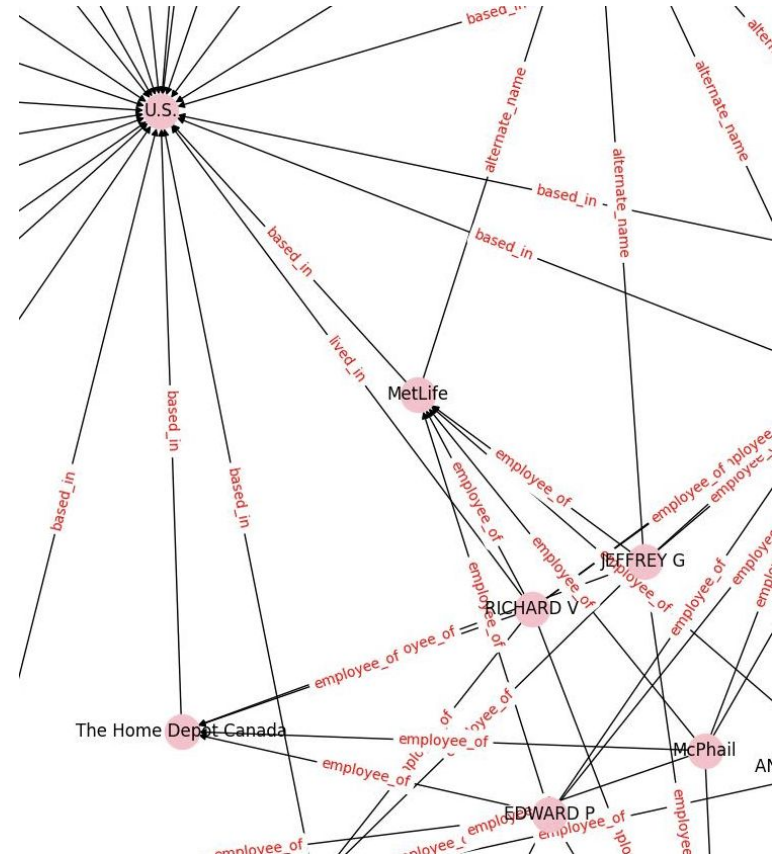
- Combination of rule-based and deep learning models for NER⁴
- Pre-trained models for various NER tasks
- Uses tokenized sentences with labeled entities to identify relationships between entities



Relation extraction. Papers With Code. (n.d.). Retrieved April 17, 2023, from <https://paperswithcode.com/task/relation-extraction>

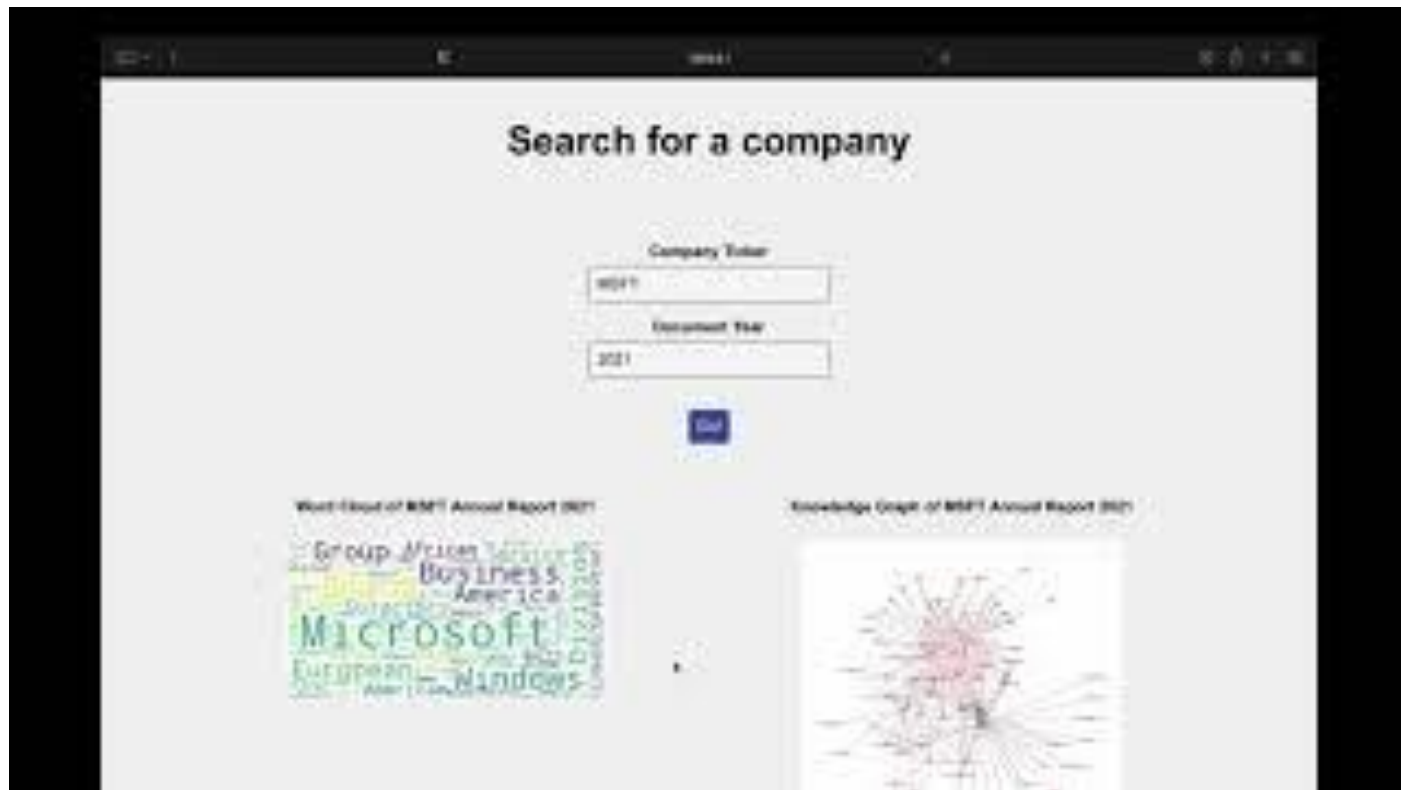
Knowledge Graph

- Combination of extracted entities and relationships
- Nodes are the entities, and the relationships are edges connecting the nodes
- Represents a more holistic view of the document



Subset of knowledge graph of Home Depot (\$HD) in 2021

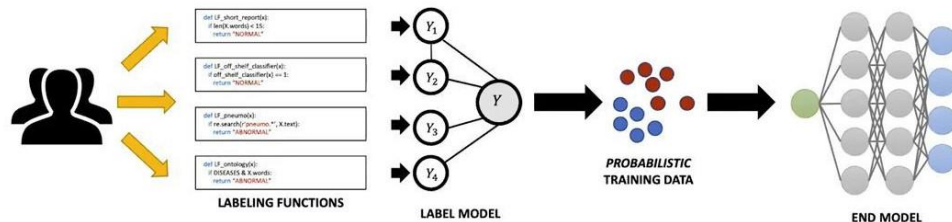
Demo



The screenshot shows a web application interface with a dark header bar. The main content area has a light gray background and features the heading "Search for a company". Below this heading are two input fields: "Company Ticker" with the value "MSFT" and "Reported Year" with the value "2021". A blue "Get" button is positioned below the input fields. At the bottom of the interface, there are two panels. The left panel, titled "Word Cloud of MSFT Annual Report 2021", displays a word cloud where "Microsoft" is the most prominent word, followed by "Business", "America", "Group", "Division", "Windows", and "Europe". The right panel, titled "Knowledge Graph of MSFT Annual Report 2021", shows a complex network graph with numerous nodes and edges, representing the relationships between entities mentioned in the report.

Looking Forward: Using FiNER, Developing FinReer

- FiNER⁵
 - Specific NER model designed for financial text
- FinREER
 - Developing weak supervision relationship pipeline on top of the existing FiNER



1. Users write **labeling functions** to create noisy labels

2. We **model** and **combine** these labels

3. The generated labels are used to **train a downstream model**

Weak supervision modeling, explained.
KDnuggets. (n.d.). Retrieved April 17, 2023,
from
<https://www.kdnuggets.com/2022/05/weak-supervision-modeling-explained.html>



Looking Forward: Deploying Website, Limitations

- Results are extremely variable depending on wording on 10-Ks
- Generated knowledge graphs are very large
 - Difficult to visualize
 - Tradeoff between including details and keeping only important parts
 - Leads into questions about how to determine importance of relationships and/or entities
- NER and relationship recognition are slow, resource intensive
 - Website cannot be made faster, will become only slower as more work is done
 - Website doesn't really have much scalability in current state

Looking Forward: Conclusion



Bridges the information gap by providing an easy, accessible way to obtain key information contained in 10-K filings



Continue improving our model by developing custom relationship extraction models specific to the finance domain



Deploy website for public use and continue adding new features to deepen document analysis



References

- ¹ Machokoto, M., Sikochi, A., & Harris, T. (2022). Do peer firms influence annual report readability? *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.4308253>
- ² Stanford University. (n.d.). *Stock prediction using Twitter sentiment analysis*. Retrieved April 18, 2023, from
<https://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf>
- ³ Cao, S., Jiang, W., Yang, B., & Zhang, A. (2020). How to talk when a machine is listening?: Corporate disclosure in the age of ai.
<https://doi.org/10.3386/w27950>
- ⁴ Flair. (n.d.). *Welcome to Flair's documentation!* Flair documentation. Retrieved April 17, 2023, from
<https://flair.readthedocs.io/en/latest/index.html>
- ⁵ Shah, A., Vithani, R., Gullapalli, A., & Chava, S. (2023). FiNER: Financial Named Entity Recognition Dataset and Weak-Supervision Model.
arXiv preprint arXiv:2302.11157.



Questions?