Release Management Extraction Automation

Hridhay Monangi, Rohit Ramkumar, Vishruth Guruvayurappan, Rohan Potta, Jacob Clark, Aakarsh Jain, Paretosha Singh, Maanav Kyabarsi, Arvind Ramadhas Gomathy, Ivan Ni, Yilan Zhang

celito

Introduction

About Celito:
- Celito is a consulting firm with many clients in the life sciences industry
- These clients need to stay compliant with GxP regulations, requiring that any software updates be evaluated and documented to ensure companies produce safe, quality, products
- As a result, Celito makes manual impact assessment documents of various document software for this clients to help them maintain compliance
- As software updates become more frequent and complex, Celito wants to automate the impact assessment process for efficiency and error reduction

Project Overview:
1. Automatically extract release note content periodically
2. Analyze impact levels using an NLP model
3. Adjust impact levels according to client requirements
4. Generate a comprehensive impact assessment document that is ready to send to client

celito

RELEASE IMPACT ASSESSMENT	
TL-CTI-016 Ver. 1.0	Page No. 1
Parent Doc ID. SOP-CTI-017	
Title: Impact Assessment for DocuSign Part 11 Release July 2024	

Feature Name	Feature Description	Predicted Impact
Announcements	Learn about the announcements for the latest release ofDocusignCLM. Please review the following announcements for the February 2025 release.	No_Impact
Release Windows	The release windows are as follows: UAT release isThursday, February 6, 2025, from 5:00 PM to 11:00 PM, U.S. CentralTime. No downtime is expected during the release window. Production release isFriday, February 21, 2025, from 10:00 AM to 3:30 PM, U.SCentral Time for EU data centers, and 5:00 PM to 11:00 PM, U.S. Central Time forNA/US data centers. No downtime is expected during the release windows.	No_Impact
Invoke Web Service Step No Longer	As of this release, theInvoke Web	No_Impact

The document above is an impact assessment document generated by our model, capturing feature names, descriptions, and impact levels
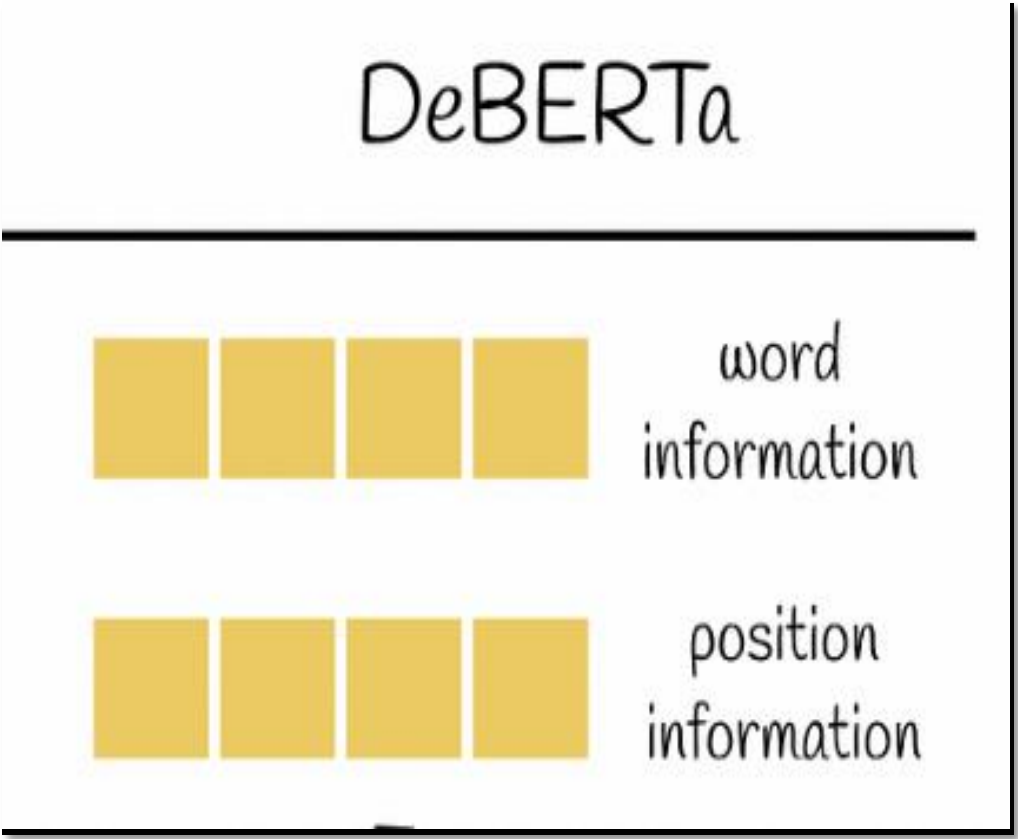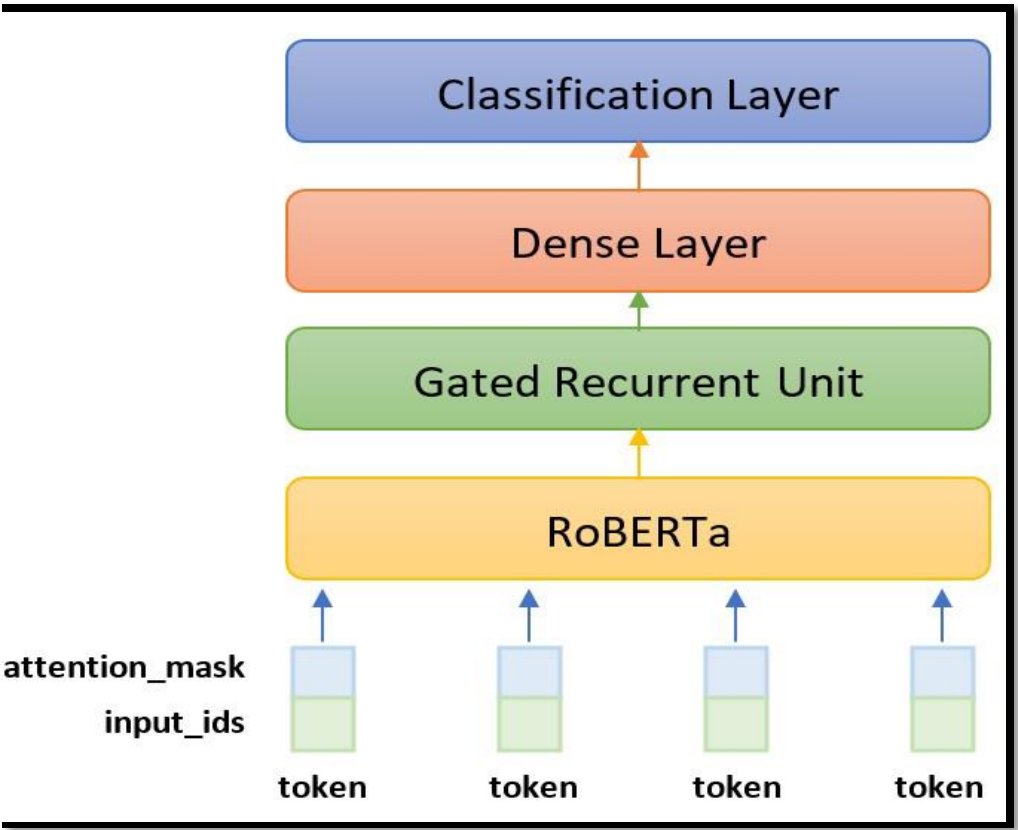
2. IMPACT ASSESSMENT

- We started by manually preprocessing completed impact assessments from software companies like DocuSign, Veeva, and ZenQMS, using them as training data for our model

- To enhance our model, we generated synthetic impact assessment data to expand the variety of our training set, helping improve classification accuracy

- We utilized the RoBERTa NLP model to analyze the significance of each update and classify updates into impact categories (None, Low, Medium, High)

- We complete our analysis by using a cosine similarity matrix to compare features with client requirements to adjust impact levels depending on similarity

- The final result is a word documents of our impact assessment, as shown in the figure on the left
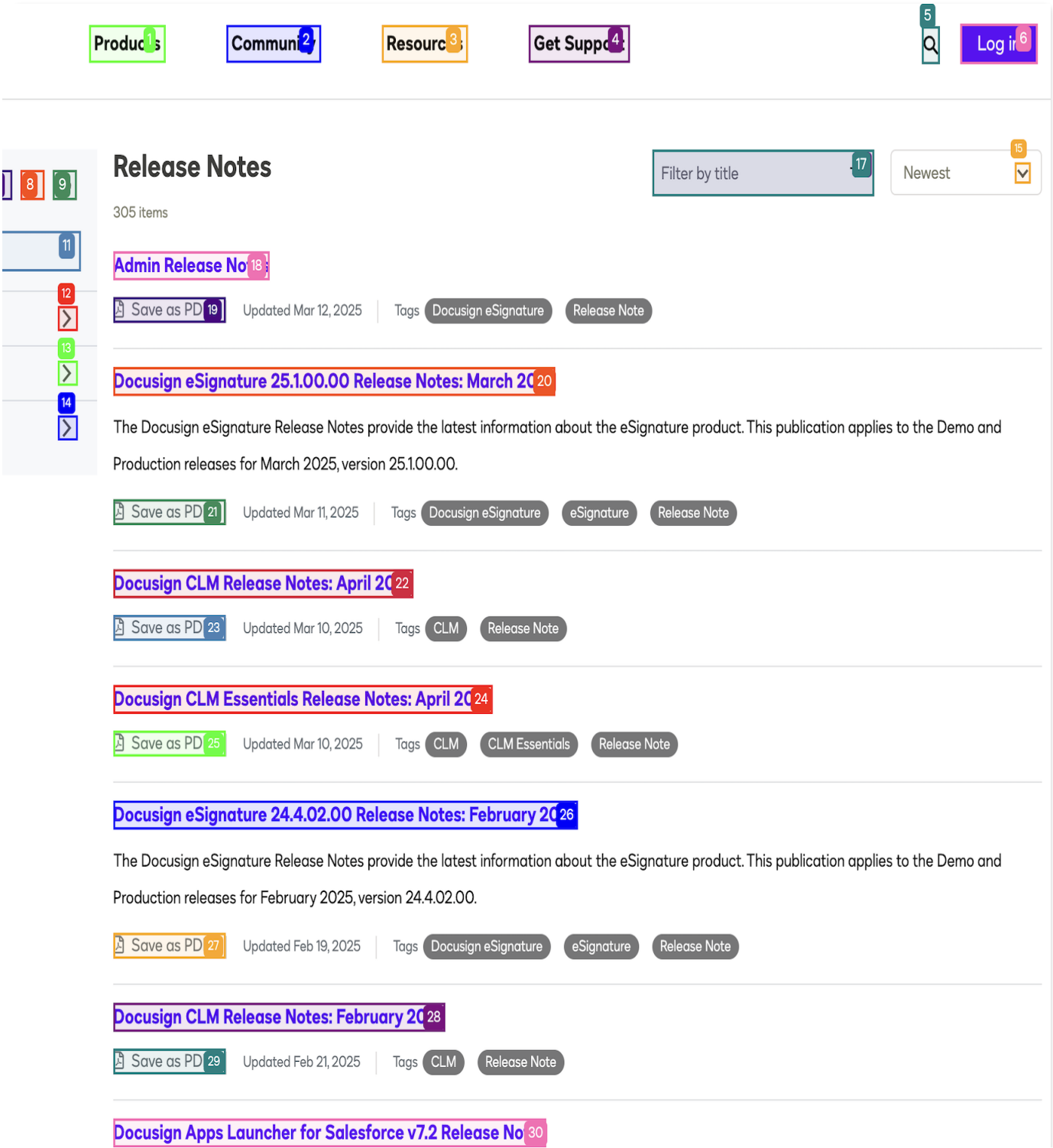
An example of Browser Use agent in action, capturing release note information automatically

1. EXTRACTION

- We start by using an AI Agent called Browser Use to navigate to each website and extract the most recent release note

- Browser Use classifies and sorts through different elements of a webpage, as can be seen in the figure on the left, and based on this information can browse the webpage and accomplish a goal set by a user prompt (like a ChatGPT prompt)

- We prompted Browser Use to navigate to the different websites we want to monitor and return the URL and raw HTML code of the most recent release note published

- We then use a model to extract and format the content/explanation of the release based off the raw HTML code

- A DeBERTa NLP model is then used to classify the contents of the update into titles, headers and content, which will be formatted into a tsv file to serve as the input for the impact assessment model

Classification Layer

Dense Layer

Gated Recurrent Unit

RoBERTa

attention_mask
input_ids

token   token   token   token

DeBERTa

word information

position information

Conclusions and Future Goals

- Browser Use efficiently extracts the latest release notes from multiple websites, ensuring timely data collection

- RoBERTa model improves classification accuracy for the impact assessments and automates report generation

- The solution supports multiple clients, reducing manual labor and increasing operational efficiency

- Future goals include refining NLP models, expanding training data for both extraction and impact assessments to improve model accuracy

Acknowledgements