

LORE_{AI}

devex

How Devex used Salient to reduce costs and become more efficient.

Background

Devex is a media platform for the global development community. They analyze thousands of their global partners' documents including annual reports, end of project summary and results reports, impact reports, general studies, and financial reports. With such a high volume of documents, it is difficult for Devex to comb through and extract valuable and granular information. If tagged at all, the documents will usually only have high level tags such as country, sector and operating periods. As a result, Devex analysts spend an enormous amount of time manually reading through these reports to identify and tag precise information.



Previous Approach

Before Salient, Devex's pursued both a manual and an automated approach to identify and label documents by project location. Initially, analysts opened documents (mostly PDF's) one by one, searched for a specific sentence or phrase that indicated the location of the project using the PDF search tool, and then copied and pasted that location and associated document identifier into a spreadsheet. Analysts were unable to simply use keyword filters to search the data because the search was either insufficient or cast too wide a net. In addition, the analyst's subject matter expertise was critical because the language used in the reports varied significantly.

Due to the tedious and time-consuming nature of the process, Devex attempted to automate this processes using a well known machine learning service. Using a dataset of labeled PDFs generated laboriously by their analysts they trained the system but they found it lacked the granularity to correctly label project location at a sub-national level.

The Solution

Devex uploaded several thousand documents into Salient for analysis and to automate their process.

Devex analysts used Salient's **Semantic Search** to access all of their documents at once, rather than one at a time, and use the tagged metadata filters to narrow their search. They trained **Smart Highlighters** interactively to label samples of phrases specifying *relevant* locations. During this process Salient proposed new suggestions as it learned, saving the analysts a huge amount of time in finding and labeling examples. Once the highlighters were trained they analyzed all the documents (millions of sentences) at a rate of ~10,000 sentences per second looking for anything relevant for identifying project locations.

By identifying and tagging only relevant sentences, Salient was able to accurately extract project locations from the reports at a sub-national level. The analysts exported Salient's results and compared them to their existing automated solution. Salient correctly picked out only relevant locations within the country project (upper image) while the other machine learning solution simply extracted any location mentioned in the document (bottom image).

Previous Automated Approach



Salient Approach



Results

Using Salient's combination of integrated semantic search, metadata filters, and highlighters, analysts were able to match and extract granular location information from thousands of documents in a fraction of the time they would have spent on the task otherwise. Devex also saved money that would have been spent on outside contractors to perform the same task. In addition to the cost and time savings, Devex was able to train and produce multiple Smart Highlighters to find specific information instantly. These Smart Highlighters are now embedded in the Devex team and can be applied to any new documents at any time, without any extra work by the analysts.

Salient saved Devex significant resources and provided a powerful, accurate and scalable solution.

Devex is now pursuing funding to scale this analysis up to several hundred thousand documents.

***Smart Highlighters** are a powerful classification and extraction tool built into Salient. They integrate seamlessly into an analyst's workflow and are used to highlight sentences of interest in any document. After a few examples, Salient learns to automatically highlight other similar sentences based on semantics instead of keywords. Through a simple training loop, the analysts makes corrections to the AI and it quickly adapts to those changes, making it even more effective. Once trained, highlighters can be scaled and reused immediately.*