

# Text Summarization: Our Approach

Finch for Text® is powerful natural language processing software that enables unique, accurate and thorough understanding of content – in the form of news, notes, reports, message traffic, transcripts or any other kind of unstructured text. Below is a sampling of how we apply text summarization in support of a number of business and mission critical use cases.

## What is Text Summarization?

Text Summarization involves creating short, fluent summaries of documents without losing key information or the document's overall meaning and intent. Doing this accurately is challenging – doing it accurately, quickly and at scale proves even more so.

Presently, there are two main ways of approaching text summarization. The first is called extractive summarization, which involves selecting a subset of phrases or sentences in the original text and combining them to form a summary. The second type, abstractive summarization, involves generating phrases or sentences, possibly rephrasing them or using words that were not in the original text. Abstractive summarization provides more concise and human-like summaries but can be slow and inaccurate on longer documents. Additionally, redundancy becomes an issue when generating multi-sentence summaries.

---

## Our Approach

At Finch Computing, we use both abstractive and extractive summarization to produce better, more accurate summaries for our customers. We leverage cutting-edge summarization models that combine both approaches and use reinforcement learning to ensure the two approaches work in harmony.

We've trained those hybrid models on *more than five years* of news content. This intensive training on a large and broad corpus of text enables us to use the models on our customers' text and to deliver accurate, high-quality results that our customers can trust.

Similar to the way a human would summarize a long document, our model first uses an extractor agent to select salient sentences or highlights. Then, it employs an abstractor network to rewrite (i.e., compress and paraphrase) each of these extracted sentences.

Using our experience in this domain and proprietary real-time computing approaches, we are able to manage text summarization projects for customers that involve assessing huge volumes of text – and to do it quickly and accurately.

# Text Summarization: At Work for Customers

## Multiple Use Cases

Our text summarization models can be used to summarize documents ranging from:



Summarizing news documents to identify trends



Analyzing engineering reports to mitigate risk



Assessing internal emails to identify insider threats or to gauge employee opinions



Monitoring message boards and chat rooms



Identifying opportunities for cost savings and efficiencies

## An Example

Below is a comparison of how human summarization compares to our text extraction approach on a piece of raw news text. As is shown, the Finch summarization is right in line with how a human would summarize the piece. Finch Computing's text summarization approach offers customer better, more accurate summaries that capture the correct, salient points and preserve the overall meaning and intent of the documents.

"four workers died in a massive oil rig fire that raged for hours off the coast of mexico wednesday. mexican state oil company pemex said 45 workers were injured in the blaze, which began early wednesday morning. two of them are in serious condition, the company said. authorities evacuated about 300 people from the abkatun permanente platform after the fire started, pemex said. at least 10 boats worked to battle the blaze for hours. the fire had been extinguished by wednesday night, pemex said in a twitter post. the company denied rumors that the platform had collapsed and said there was no oil spill as a result of the fire. the state oil company has n't said what caused the fire on the platform, which is located in the gulf of mexico's campeche sound. the fire began in the platform's dehydration and pumping area, pemex said. cnn's mayra cuevas contributed to this report."

### Human Summary:

- the fire on a platform in the gulf of mexico **has been extinguished**, pemex says.
- **45 workers were injured in the blaze**, according to the state oil company.
- **four workers were killed in the oil rig fire**, which started early wednesday.

### Finch Summary:

- **Four workers died in a massive oil rig fire.**
- **45 workers were injured in the blaze.**
- Authorities evacuated 300 people from the Abkatun Permanente platform..
- The fire **had been extinguished** by Wednesday night.