
Coretraq Technology Whitepaper

21st March 2017

Introduction

Contact Centers are growing but Small and Medium Businesses and well as Enterprise Business Units need effective custom vocabulary driven Audio Mining / Speech Analytics products that are cost effective.

Accurate B2B speech systems need to understand product and brand names, as well as industry and company specific business terms; this requires the creation of custom speech vocabularies unique to each customer. At Yactraq we focus on automating the creation of custom speech vocabularies with three distinct advantages over legacy solutions: speed, scale, and cost. We can create custom speech vocabularies in a matter of days that consist of hundreds of thousands of words, and all at a fraction of the cost of legacy solutions.

What enables Yactraq's technical capabilities is Coretraq, Yactraq's proprietary and patent pending speech based semantic platform. Coretraq's custom vocabulary capability is available as an API or via a user interface (OmniTraq).

Additionally, Yactraq is an IP-For-Defense partner of Intellectual Ventures, which uniquely poises Yactraq in the Speech Tech industry and provides Yactraq with an unparalleled freedom to operate and bring disruptive solutions to the market.

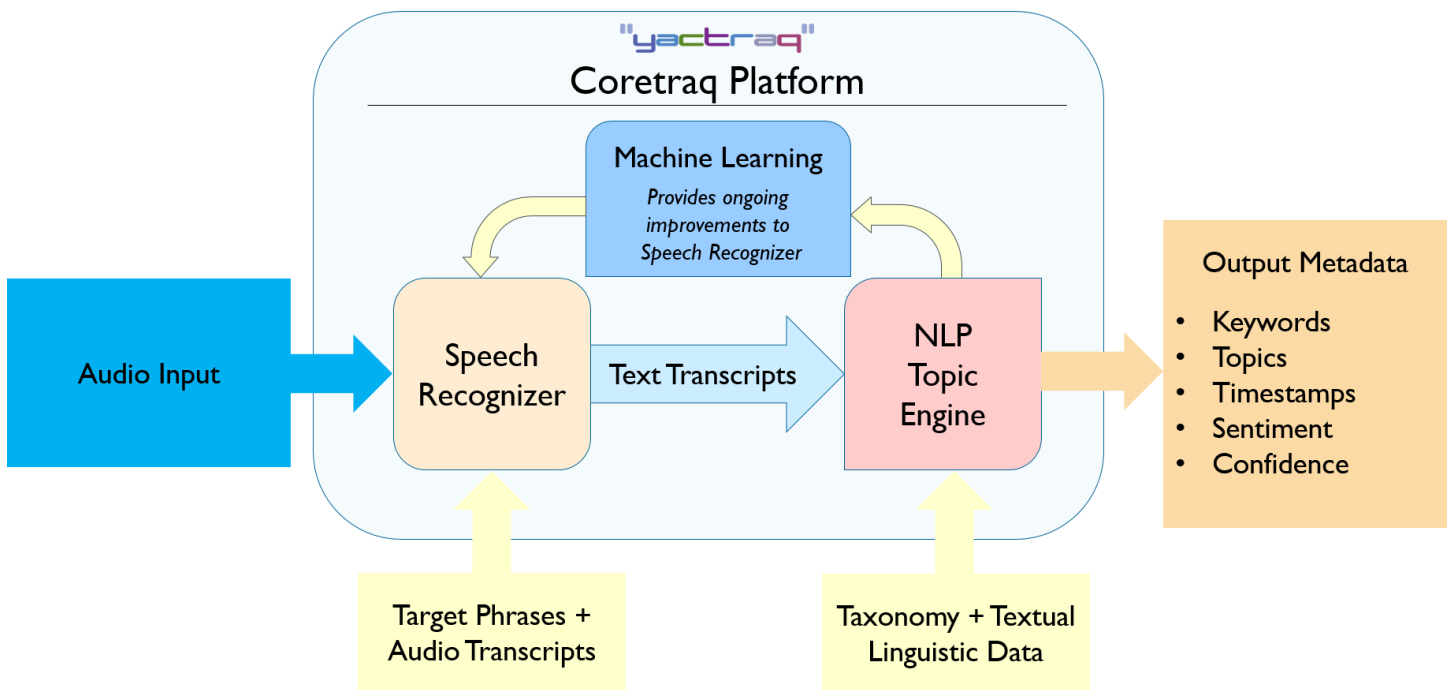
Present Status

Yactraq's key technical approach to deliver business value to customers is through LVCSR technology combined with custom vocabularies, built quickly and inexpensively through increasing degrees of automation; we call this our Coretraq platform.

We have built custom vocabularies with taxonomies like Open Directory, as well as customer specific target taxonomies & lists. Taxonomic data is first used to generate search terms, then standard web crawling techniques and API's are used to collect web pages of related linguistic data. The taxonomic and linguistic data is then provided as system data to Yactraq's NLU (Natural Language Understanding) module. Data from the NLU module is then exported out and further provided as input to a machine learning process. The output of Yactraq's machine learning process is provided as input to a set of tools which generate a statistical language model for use by Yactraq's speech engines. Yactraq also uses audio transcripts in its machine training process.

At run time, audio data is decoded and sent to a pre-processor; the pre-processor then eliminates music, noise, silence from the data and improves the signal to noise ratio. Only speech segments are sent on further to the speech recognizer. The outcome is higher accuracy because 8khz audio is often noisy, as well as throughput gains of 100X because media based operational data often contains substantial sections of music or noise. The output text from the speech recognizer is further processed via Yactraq's proprietary machine learning and natural language understanding techniques by Yactraq's NLU module. This additional processing allows our platform to quantitatively determine the primary subject topics of the given minute of audio or video data.

The current version of Coretraq has been proven across approximately five million minutes of phone calls and videos, and delivers high levels of metadata accuracy. This validates that Yactraq’s patent pending process as described above can generate custom vocabularies that are orders of magnitude faster and cheaper than legacy approaches.



The yellow objects above represent machine training data; the blue boxes represent operational run-time customer data.

In specific customer cases, Yactraq call analytics customers have seen around 99% key phrase spotting accuracy; and have auto generated semantic signatures for web videos across 20,000 leaf level topics structured in a 4-layer deep open directory taxonomy using a linguistic training data containing more than 10 million sentences with a lexicon of 300,000 words. This linguistic data is then further compressed into a statistical language model for use various Yactraq speech recognizers.

Yactraq call center customers have seen 400% productivity gains for managers and quality analysts; and a key Yactraq video search client was acquired by a top 5 US digital publisher, proving the commercial value of Yactraq’s Coretraq platform and its custom vocabulary capability.

Patent Pending Fusion of NLU and Speech Recognition

Our pending patent is a method for automating the generation of language models (LM’s) in speech recognition systems. As indicated above, this process uses the NLU module to trigger a set of machine learning steps resulting in the automated generation of a speech language model. The key business benefit of this method is that it provides multiple order of magnitude improvements in the speed and cost of generating custom vocabulary LM’s.

Coretraq - Highlights

Versatile Speech Recognition:

The information compression capability provided by Yactraq's proprietary machine learning tools is a key asset in building embedded systems as it allows large topic sets and lexicon's to be compressed, with minimal loss of information, into smaller footprint embeddable speech recognizers.

Yactraq has already processed 5 million minutes of customer data using Yactraq's speech recognizers, which have proven their versatility across scalable cloud environments as well as mobile and embedded environments. Additionally, Yactraq speech scientists have world class speech capabilities. Their past work has found application in the Eurofighter cockpit, Samsung mobile phones, and even watches.

Deep Neural Networks:

Yactraq speech engines employ Deep Neural Network (DNN) technology. Acoustic modeling is an area where neural network based systems have shown great promise in the last few years and are a key aspect of Yactraq's approach to dealing with:

- Multiple languages
- Multiple accents
- Acoustically challenging data

Class Based Language Models:

Crawling web pages was sufficient to build a video search capability because essentially the problem that video search systems try to solve is - determining the degree of similarity between data object A and data object B. But business intelligence applications have deeper requirements driven by the need to answer more specific questions.

For example, a call analytics Voice-Of-The-Customer application may want to know why the same issue sometimes can be resolved on a single customer service phone call, but can also take more than five calls on a separate occasion. In such case the system needs to determine negative sentiment associated with very specific business objects. Also in such cases, only limited amounts of linguistic data may be available. The language patterns required may not be available through crawling of web pages, and in such cases class based language models can be used very effectively.