



PodMining Indexer

Overview

If you are looking for a high end speech recognition product that can automate the process of indexing podcast feeds, your search ends here.

With the SAIL LABS PodMining Indexer you can process audio files in various formats and produce annotated text output that can be searched for keywords and topics.

The advanced speech technologies can inherently handle variations in speaking styles as well as recording environments for different domains.

Distributed architecture and built-in Application Programming Interfaces facilitate integration with your existing infrastructure.

Time tagging of each word helps to rapidly identify and access the target segments that interest you. Podcast search is made easier for you by identification of key speakers and categorization of raw audio into specific topics.

The PodMining Indexer serves as an extremely cost effective solution for podcast search engines, media observation and open source intelligence requirements.

Features and Benefits

- Best-of-breed speech recognition technology enables accurate conversion of speech to text, even with high levels of background noise
- Speaker Change Detection segregates audio segments based on speaker for focused information retrieval
- Speaker Identification is your answer to "who" has been saying "what" in which podcasts
- Topic Detection helps locate stories based on your favorite topics with ease
- Named Entity Detection makes finding of relevant information simpler as words

belonging to categories like persons, locations, organizations etc are highlighted within the text output

- Scalable Architecture supports single machine to multi machine configurations, providing a cost effective solution for small and medium sized podcast search engines, media monitoring companies etc.
- Industry standard XML output facilitates integration with complementary technologies

Technical Specifications

Hardware prerequisites

The minimum hardware configuration to run PodMining Indexer is 2GB of memory and a single Pentium-III/733 processor (will run approximately 2 times slower than real-time).

PodMining Indexer is designed to perform in real time or faster with common off-the-shelf PC hardware. We support hardware running Windows XP or Windows Server 2003 on Intel or AMD CPUs; typically a Pentium-4 3.4 GHz Hyper-Threading CPU with 2GB of memory.

The PodMining Indexer requires a sound card if you feed input other than PCM audio files.

Configuration hints for real-time indexing:

- Pentium 4 with hyper-threading 3 GHz or better
- 2GB ECC DDR RAM compatible with your system
- Disk space: 60 GB recommended for production use on indexers; more for machines acting as DMC; Minimum: 30GB (between 350 and 500 MB per Language Feature plus temporary disk space)

Software prerequisites

- Microsoft Windows XP SP1 or Windows 2003 Server
- DirectX 9.0 or later. This is required for indexing mpeg, mp3, wmv, avi files using the optional Media Feeder program.
- ActivePerl-5.8.7.813 from ActiveState. This can be downloaded from <http://ftp.activestate.com/ActivePerl/Windows/5.8/ActivePerl-5.8.7.813-MSWin32-x86-148120.msi>. Note: Only the mentioned version of ActivePerl is guaranteed to work with PodMining Indexer.
- Real Player (current version); this optional component should be installed if you need to index Real Media files; this can be downloaded from www.real.com

Language Options

Languages supported by PodMining Indexer are:

- Arabic
- English
- French
- German
- Spanish

Integration

The open architecture of the PodMining Indexer enables easy integration with complementary technologies for diverse applications. The integration can be done at two levels:

- Using command line tools provided by Sail Labs
- By linking with the API C++ library and directly accessing Indexer components

Technologies

PodMining harnesses the synergies of some of the best speech processing and language technologies produced or currently in development. Technologies such as Automatic Speech Recognition, Speaker Identification, Name Spotting, Topic Classification, and Story Segmentation have been integrated in our system, and together produce comprehensively indexed text files from audio input.

Automatic Speech Recognition

Automatic Speech recognition is performed in a sequence of steps; it first processes the incoming audio, then chunks the audio into sections of speech and non-speech, and then applies speech-recognition to those segments identified as containing speech. This can be done in real-time, for large vocabularies (64K entries) and for 8 and 16kHz (but is not limited to these).

Our speech recognition engine is language independent (modulo changes on acoustic front end, e.g. for tonal languages). Languages the recognizer has been run in include: English, French, Spanish, German and Arabic.

Front-End: Standard MFCC coefficients, energy+ 5/3 frame regression, deltas, delta-deltas, cepstral mean subtraction.

Acoustic Models: Speaker- and Gender-independent acoustic models, 3- and 5-phone context models, with Gaussian Prototypes tied at the prototype as well as mixture weight levels.

Language Models: Our engine uses mainly Bigrams and Trigrams for the Language Model. We adopt a Witten-Bell type back-off model. The Language Model and Acoustic scores are combined for maximum accuracy.

Decoder Search: We use a multi-pass, time-synchronous search. During processing, increasingly detailed models are used at each step. After a forward pass and a backward pass, the resulting N-best list is re-scored using the most detailed acoustic models.

Speaker ID/Clustering

The Speaker Identification (SID) system identifies speakers in the area of broadcast news transcription. The incoming audio is first split up according to speech / non-speech regions and speaker turns are hypothesized on these chunks. Speaker clustering (SC) and Speaker ID (SID) are run on the resulting chunks. Typically from about 20 to 100 speakers can be

identified; for non-target speakers the gender is detected and the unknown speaker's segments labeled accordingly. Speaker ID/Clustering is language independent.

Speaker Identification: SID uses Gaussian Mixture Models (GMM) for a number of pre-selected target speakers and an additional number of cohort speakers who serve for normalization and gender detection purposes.

Speaker Clustering: Speaker Clustering is based on clustering pre-determined chunks (from initial segmentation) into k clusters. As quality measure the within class dispersion is used. Segments are clustered using a variant of the generalized likelihood ratio criterion.

Speaker Change Detection

Speaker Change Detection (SCD) is done using a phone-level decoder, which employs a set of broad phonetic classes of speech sounds as well as non-speech sounds. Using the information produced by the decoder (i.e. the "transcript"), the SCD system sequentially hypothesizes speaker turns at phoneme boundaries. A generalized likelihood ratio test is used to determine whether a change should be made.

SCD works on the cepstral features and energy; no derivatives are used. The acoustic models employed are tied on the phoneme level with a language model based on the phoneme classes.

Story Detection

Story Detection consists of several phases. First an episode is partitioned into sections of homogeneous topics. The boundaries of these sections are adjusted further in a subsequent step. Finally, these stable sections are scored against a statistical model representing the individual topics. The top ranking topics are used to determine what the story is about.

Support Vector Machines, one model per topic and one model to model general language (all those filler words which really aren't specific to any topic). Each state represents the topic and emits topic dependent words probabilistically. Transitions between topics are allowed at every word and are not observed. At decoding time, the most likely set of topics given the recognized text is determined.

Future Prospects of Podcasting: Podvertising

Representing a shift from mass broadcasting to on-demand personalized media, podcasting will have a huge effect on the advertising industry. The future lies in automated systems that match podcasts with advertisers and user profiles by finding specific keywords within the audio material. Here, SAIL LABS PodMining represents the ideal base technology for a solution that locates target keywords and topics in podcasts and automatically inserts an advertisement that matches the content.