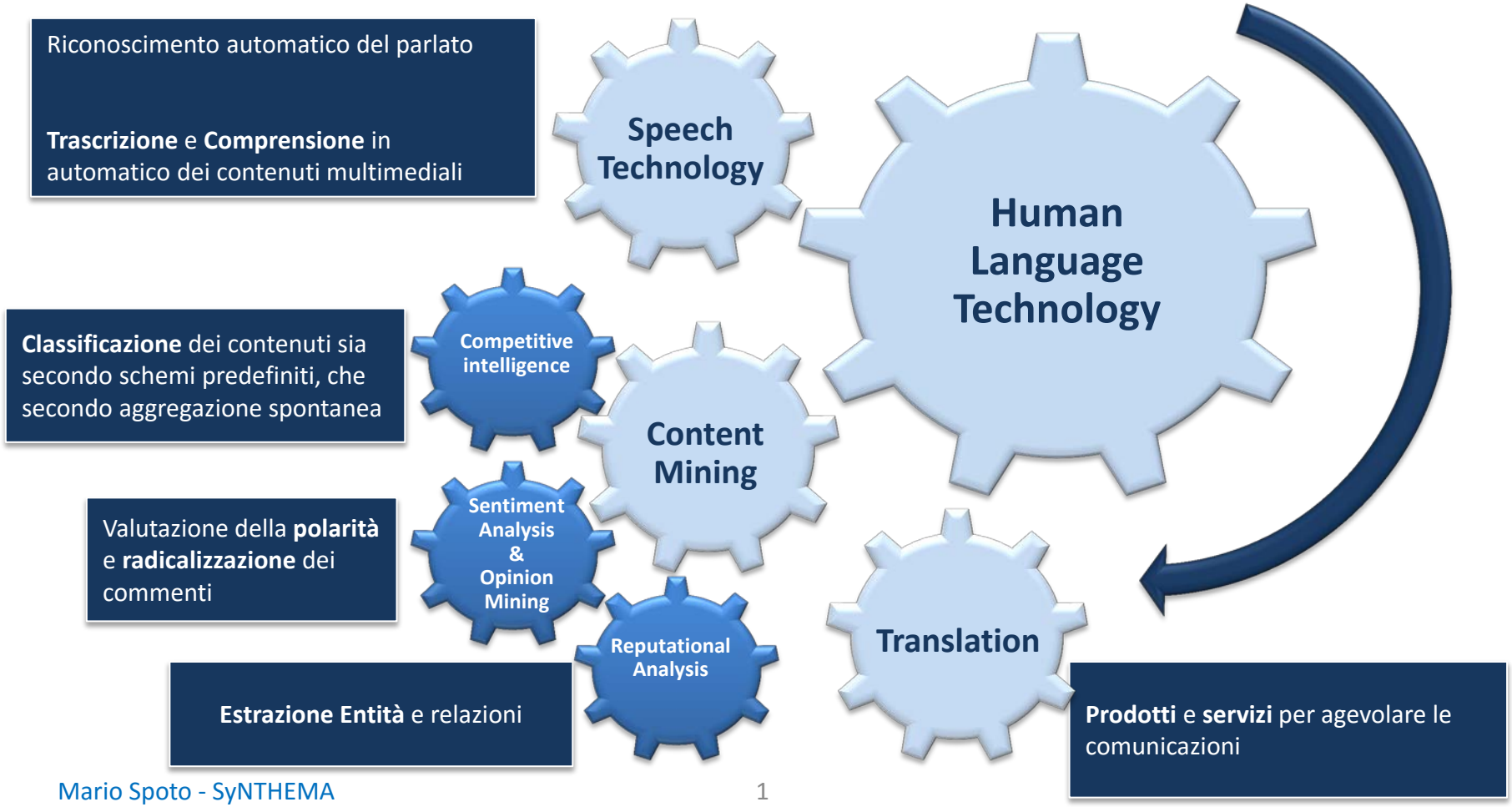
SyNTHEMA

Nasce nel 1993 a Pisa come spin-off di un Centro di Ricerca IBM.

SyNTHEMA fornisce prodotti e soluzioni per il trattamento del Linguaggio Naturale, la Traduzione e il Riconoscimento del Parlato.

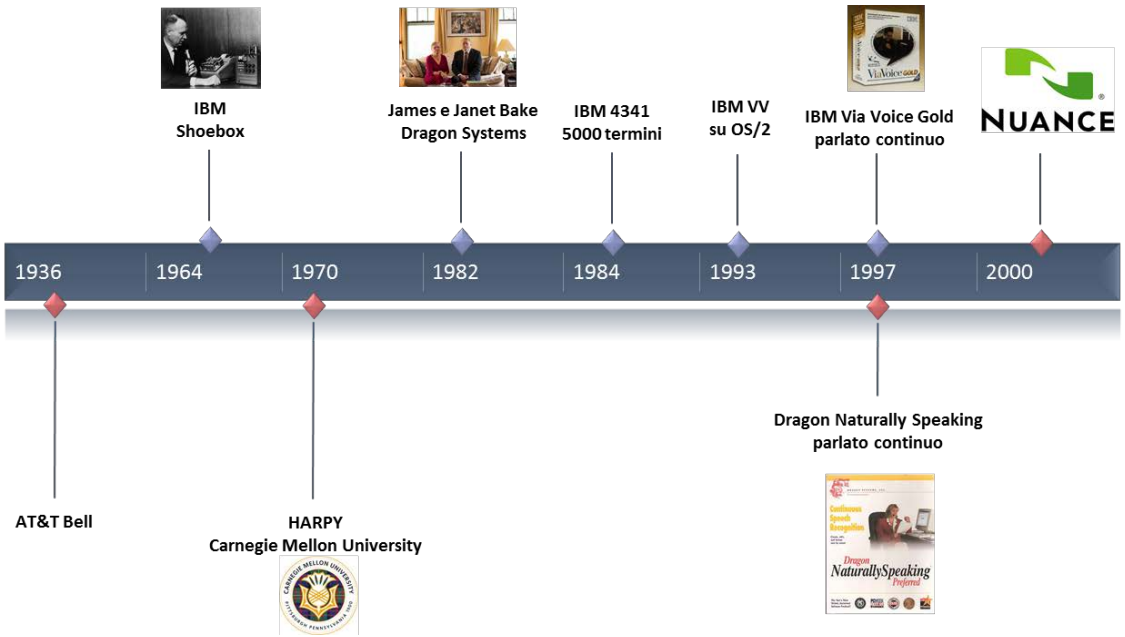


Riconoscimento vocale

La tecnologia di acquisizione vocale consente ormai svariate applicazioni e la percentuale di errore nel riconoscimento sta scendendo sempre più rapidamente. Il linguaggio parlato rappresenta senza dubbio la forma più flessibile, efficiente ed economica di comunicazione e la disponibilità di un'interfaccia vocale costituisce ormai un requisito funzionale irrinunciabile di diversi tipi di applicazioni o apparecchi elettronici.

Tra i tanti esempi, è possibile citare le ultime generazioni di telefoni cellulari o i sistemi di navigazione satellitare.

Riconoscimento vocale - Storia



I primi studi e le prime ricerche sulla possibilità di trascrivere la voce umana risalgono al **1936** e furono eseguite presso i Laboratori AT&T Bell. Ma è solo nel secondo dopoguerra che si realizzano studi e ricerche che porteranno poi a prototipi o sistemi funzionanti.

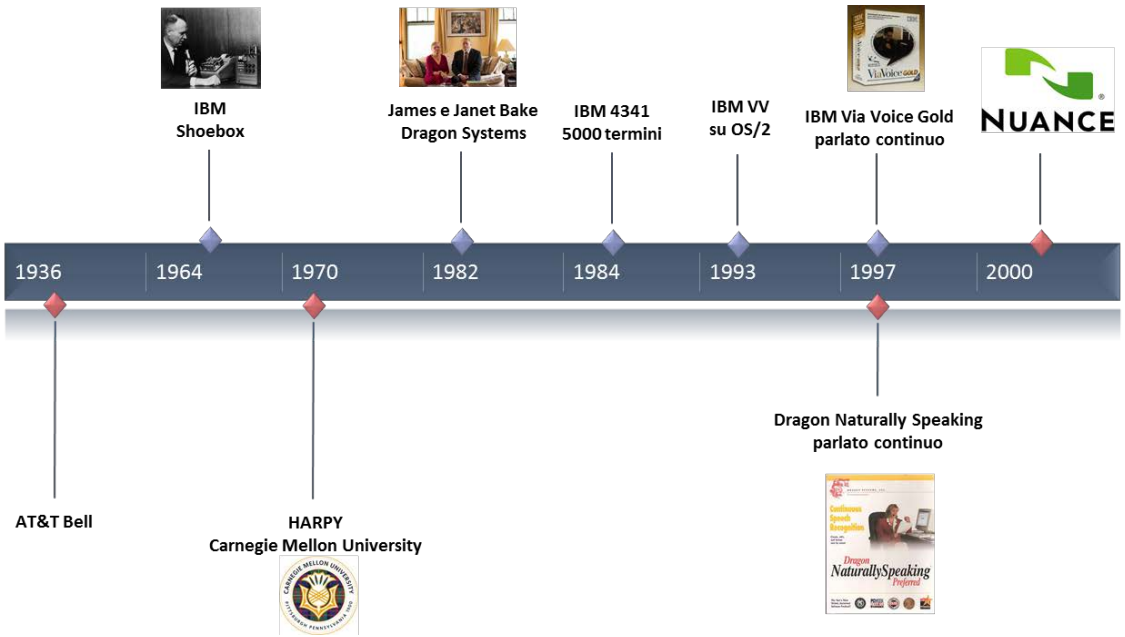
Alla **fine degli anni '50**, IBM avvia nei propri laboratori una serie di ricerche finalizzate alla realizzazione di sistemi di riconoscimento vocale. Vengono creati computer e applicazioni che permettono di riconoscere specifici pattern linguistici (parole) e di individuare una correlazione statistica tra i suoni e le parole che questi rappresentano. Nel **1964**, IBM presenta un sistema in grado di riconoscere numeri dettati dall'uomo e un vocabolario di 16 parole (IBM Shoebox*).

Negli **anni '60**, il Dipartimento della Difesa statunitense finanzia una serie di progetti che hanno come scopo la riduzione dei tempi di processing e lo sviluppo di sistemi in grado di riconoscere anche soltanto un numero limitato di oratori.

Nei primi **anni '70** nasce quindi HARPY, un sistema sviluppato alla Carnegie Mellon University. HARPY è costituito da 50 computer PDP11 funzionanti in parallelo ed è in grado di riconoscere frasi complete di senso compiuto costruite utilizzando un numero ristretto di parole predefinite **.

* https://www-03.ibm.com/ibm/history/exhibits/specialprod1/specialprod1_7.html
** https://www.youtube.com/watch?v=N3i6NoUZsSw

Riconoscimento vocale - Storia



Nel **1982** James e Janet Bake fondano Dragon Systems, che rilascia la prima versione per DOS del loro sistema di riconoscimento vocale in grado di riconoscere singole parole. Il loro primo prodotto in grado di riconoscere il parlato continuo è Dragon NaturallySpeaking del 1997 con un vocabolario di circa 23000 parole.

Nel **1984** IBM presenta il primo sistema dotato di un dizionario di 5000 termini (estesi a 20000 nel 1987). Tali sistema garantisce un'accuratezza nel riconoscimento superiore al 95% ed è supportato da mainframe 4341.
Nel **1993** IBM lancia il primo sistema di dettatura supportato da PC dotati di scheda audio e microfono convenzionale. Il sistema operativo è OS/2 e il programma garantisce una velocità di acquisizione pari a 80 parole al minuto e un'accuratezza pari al 95%. Le lingue supportate sono l'inglese, il francese, il tedesco e l'italiano.
Nel **1997** viene distribuito IBM ViaVoice Gold, che permette anche il dettato continuo.

Nel **2000**, Lernout & Hauspie compra Dragon Systems e nel 2001, Scansoft acquista da Lernout & Hauspie i diritti per i prodotti di riconoscimento vocale. Nel 2003, Scansoft acquista Speechworks e nel 2005 cambia il nome in Nuance e diventa negli ultimi anni il leader mondiale nel mercato dei prodotti commerciali di riconoscimento vocale.

Riconoscimento vocale - Tecnologia

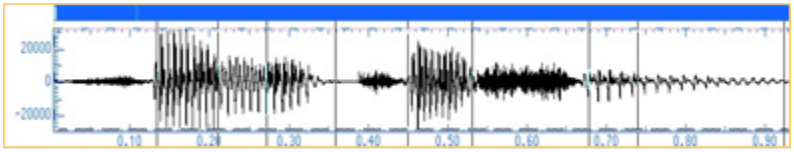
La frase viene raccolta da un microfono e la sequenza registrata dei fonemi viene confrontata con le sequenze presenti in un dizionario. Applicando modelli statistici, vengono trovate le sequenze che più si avvicinano a quelle pronunciate. Tali sequenza vengono quindi convertite nella loro rappresentazione "scritta", la parola.

Il processo si suddivide nelle seguenti fasi:

- l'acquisizione e digitalizzazione del parlato
- la sua Rappresentazione spettrale
- la Segmentazione dello spettro per l'individuazione dei fonemi
- la Ricerca nelle basi di dati lessicali di parole che possano "soddisfare" foneticamente quanto acquisito
- la Correzione su base statistica

Acquisizione e digitalizzazione

Dal punto di vista fisico, il parlato è una successione di onde di pressione (le "onde sonore") che si propagano nell'aria. La sua rappresentazione più comune è un oscillogramma, in cui nell'asse delle ascisse è rappresentato il fluire del tempo, mentre nell'asse delle ordinate viene visualizzata la variazione di pressione.



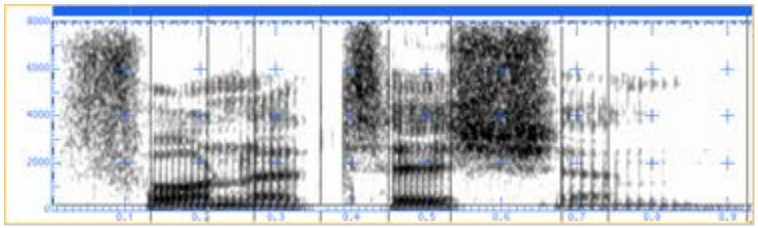
Oscillogramma della parola "phonetician"

Riconoscimento vocale - Tecnologia

Rappresentazione spettrale

L'uomo si adatta facilmente alle mutate condizioni ambientali: è infatti in grado di filtrare il rumore di fondo e di ascoltare e comprende il parlato di interlocutori con velocità di eloquio, pronuncia e intensità differenti. In altri termini, l'uomo comprende "mettendo in relazione".

L'uomo è un sistema adattivo, i sistemi di acquisizione no. Questa affermazione appare evidente dal tracciato dello spettrogramma.



Spettrogramma della parola "phonetician"

Nello spettrogramma le ascisse rappresentano il fluire del tempo, mentre le ordinate indicano la frequenza del suono. La terza dimensione, l'ampiezza, è rappresentata dall'ombra.

Segmentazione dello spettro

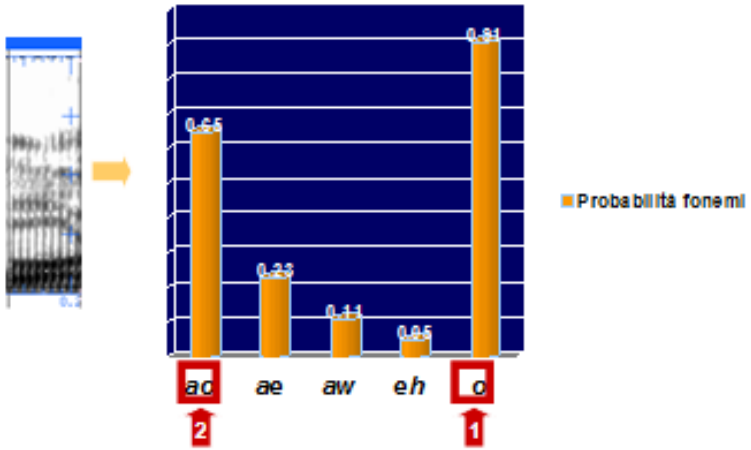
Il parlare è un fluire continuo di fonemi dove non esiste una separazione evidente tra una parola e l'altra. La segmentazione permette di ricavare dallo spettrogramma non solo i fonemi, ma anche la separazione tra le parole che questi vengono a formare.

Il processo di segmentazione è basato su algoritmi statistici noti come "catene nascoste di Markov".

Riconoscimento vocale - Tecnologia

Ricerca

Una volta "segmentati", i fonemi vengono riconosciuti su base statistica, in base alla probabilità loro associata. Esiste infatti un modello fonetico che associa ad ogni fonema una distribuzione di probabilità in uno spazio multidimensionale (ampiezza, frequenza, tempo, rumore, riverbero, ecc.).

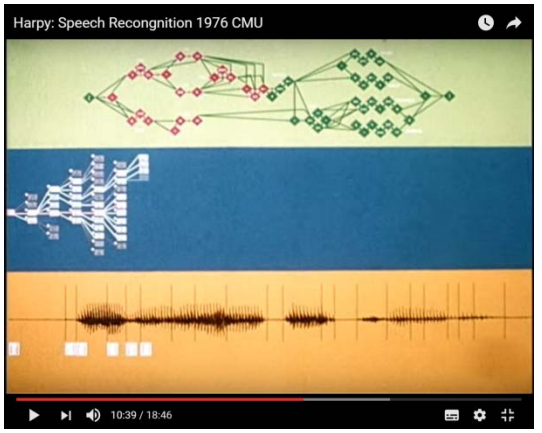


Riconoscimento di un fonema in base alla probabilità

Riconoscimento vocale - Tecnologia

Correzione

I passi precedenti producono una rete di possibili interpretazioni fonetiche, a ciascuna delle quali è associata una probabilità di accadimento. Il sistema deve cercare in una base di dati lessicali le forme che più si avvicinano alle possibili interpretazioni fonetiche. Per ogni sequenza finita di fonemi viene ricavata la lista delle parole corrispondenti più probabili. A questo punto, utilizzando un approccio statistico, è possibile scegliere la parola più adatta.



Esistono infatti grammatiche statistiche che associano una probabilità di accadimento ad ogni sequenza di tre parole. Di seguito viene descritto il meccanismo alla base di questo tipo di grammatiche:

- Vengono prese in esame le prime tre parole e le loro possibili alternative.
- Viene scelta la sequenza più probabile.
- L'analisi si sposta di una parola a destra, considerando quindi una nuova tripla composta dalle parole 2 e 3 della tripla precedente e dalla nuova. Si ripete il procedimento applicato alla tripla precedente, scegliendo la sequenza più probabile.
- Se si ottiene una sequenza altamente improbabile, con un meccanismo di backtracking si torna alla scelta precedente e si procede con una scelta alternativa.

Riconoscimento vocale - Uso

La tecnologia del Riconoscimento vocale trova applicazione:

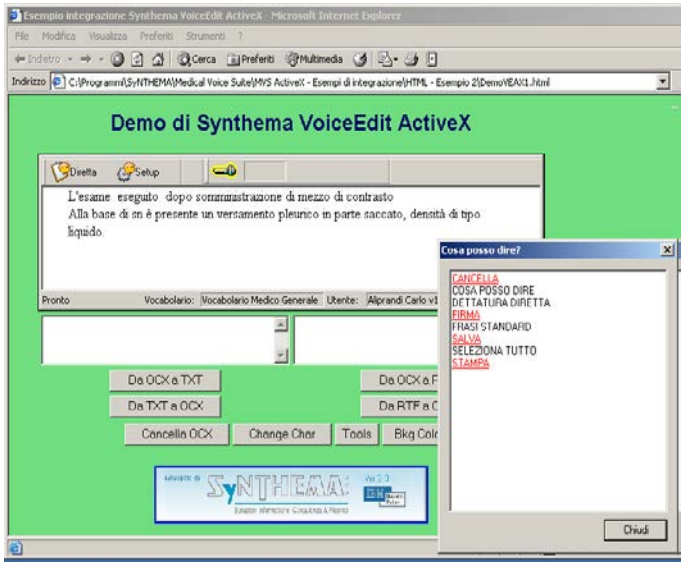
Dettatura
Composizione interattiva di un testo (Es. Refertazione medica)

Trascrizione Automatica
Trasformazione del parlato in testo (Es. Sottotitoli, Resocontazione, ecc.)

Sistemi Conversazionali
Uso della voce per interagire con un sistema (Es. CRM/Help Desk, Controllo dispositivi, ecc.)

Speech Analytics
Trascrizione di contenuti audio/video per indicizzazione e ricerca

Riconoscimento vocale – Refertazione



Sanità (pubblica e privata)

SyNTHEMA MVS - SyNTHEMA SpeechJive – DICTASPEECH (Speaker dependent)

Componente per dettatura referti medici

Command&Control

Integrabile in sistemi gestionali RIS PACS usati in ambito ospedaliero

Semplificazione attività e riduzione dei tempi di produzione del referto medico

Gestione del workflow di produzione del referto

Sincronizzazione in rete dei profili

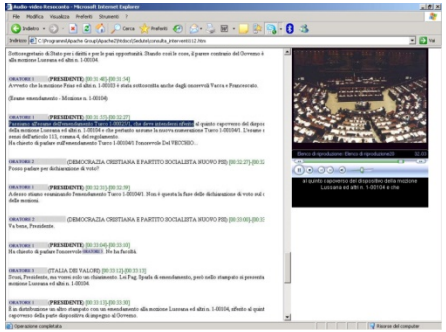
Riconoscimento vocale – Trascrizione

Media

Pubblica Amministrazione

SyNTHEMA SPEECHSCRIBE

Tecnologia innovativa di ASR Speaker Independent in grado di trascrivere automaticamente audio e video, utile per professionisti della Resocontazione e della Sottotitolazione.

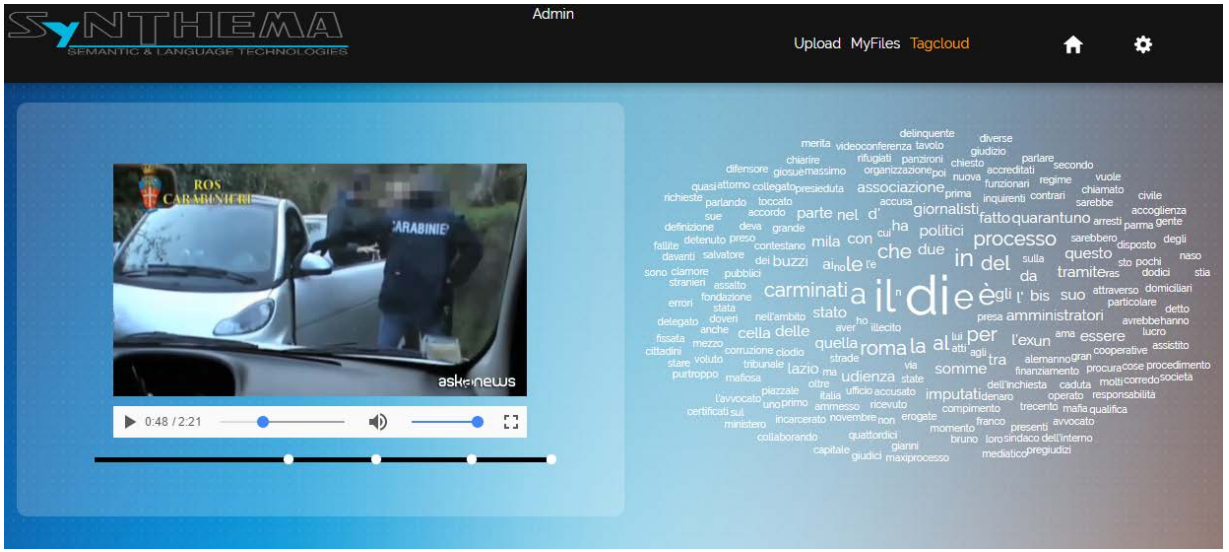


SyNTHEMA SPEECHALIGNER

Permette l'allineamento automatico tra audio e testo dei contenuti multimediali, associando univocamente alle parole il rispettivo segmento audio

Riconoscimento vocale – Conversazionali

Riconoscimento vocale – Speech Analytics



Tecnologia che permette di indicizzare (anche semanticamente) i contenuti audio/video ai fini di ricerca, creazione playlist ecc.