

moz://a

Il riconoscimento vocale nella lingua più bella del mondo

Contribuisci anche tu a Mozilla

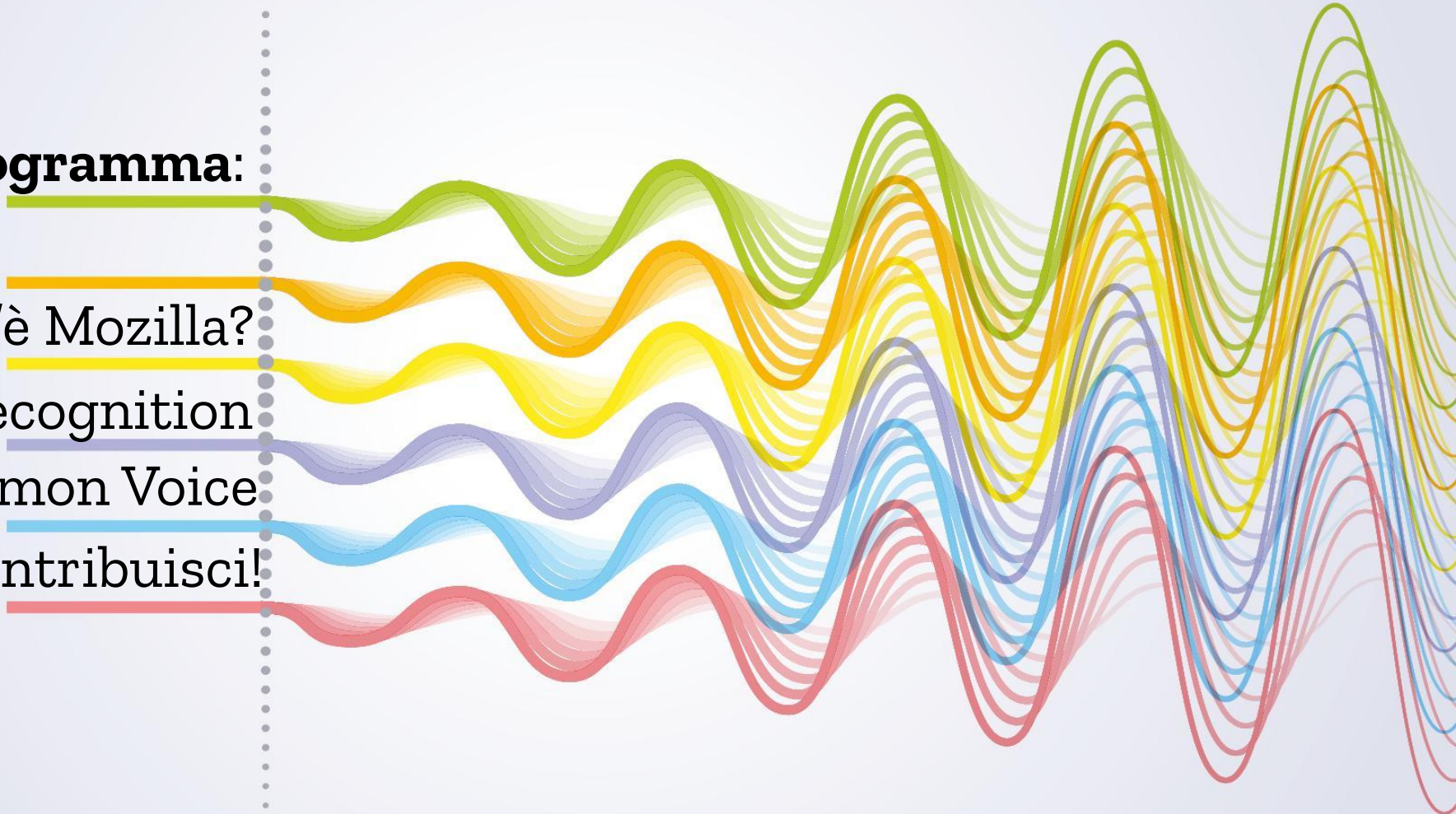
26/10/2019

Luisa Ravelli
lugBS



Programma:

- Cos'è Mozilla?
- Speech Recognition
- Common Voice
- Contribuisci!





I

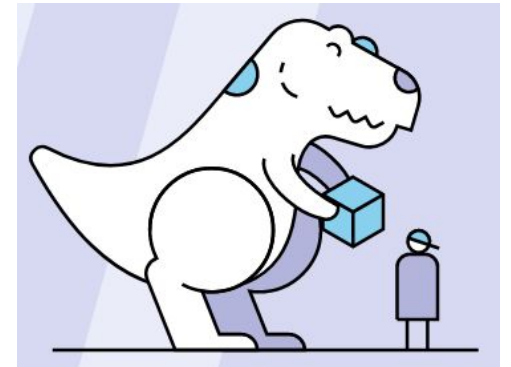
Cos'è Mozilla?



Manifesto moz://a

- Internet è una risorsa pubblica globale che deve rimanere aperta e accessibile.
- Internet deve arricchire la vita di ogni essere umano.
- L'efficacia di Internet come risorsa pubblica dipende dal suo carattere di interoperabilità (relativamente a protocolli, formati di dati, contenuto), dal suo grado di innovazione e dalla partecipazione decentralizzata a livello mondiale.
- La partecipazione commerciale allo sviluppo di Internet è in grado di apportare numerosi benefici, ma è fondamentale un equilibrio tra profitto commerciale e benefici pubblici.
- La valorizzazione degli aspetti di pubblica utilità di Internet rappresenta un obiettivo importante, che merita tempo, attenzione e impegno.

<https://www.mozilla.org/it/about/manifesto>



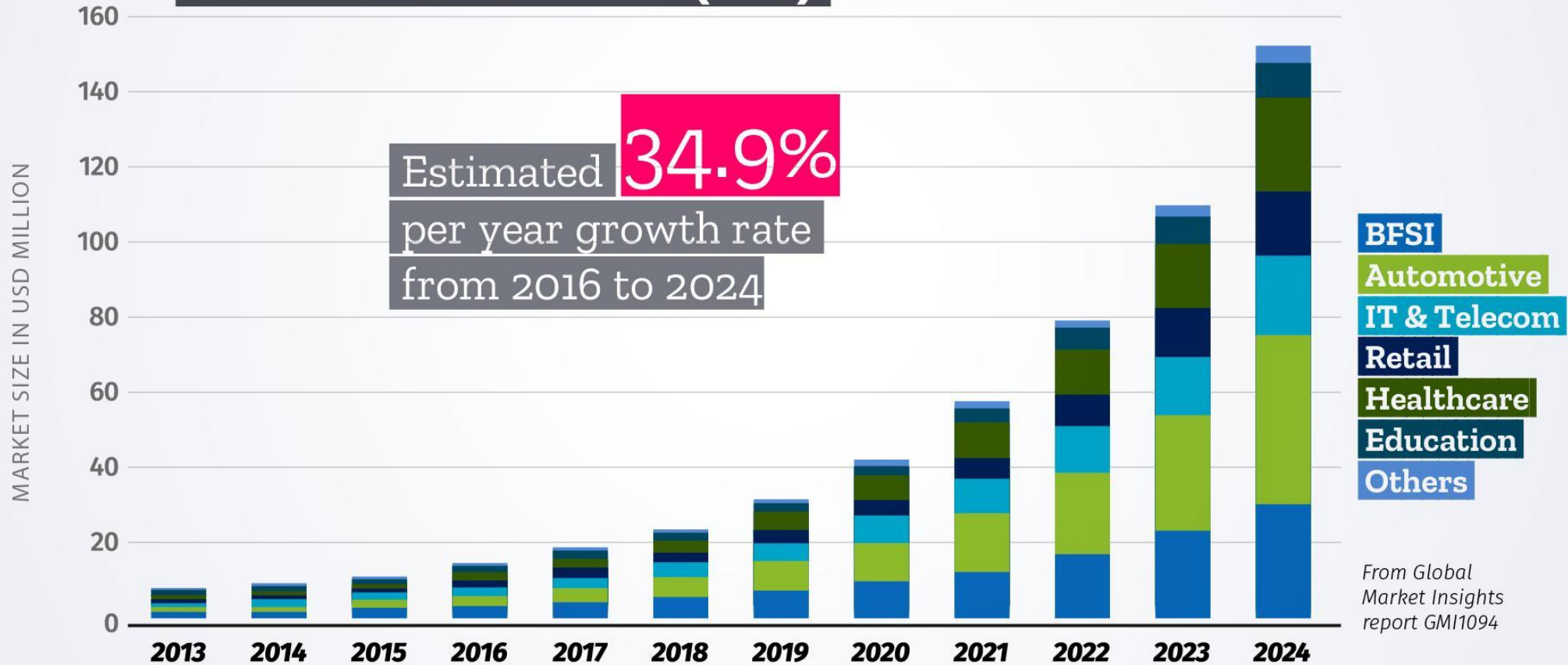
II

Speech Recognition



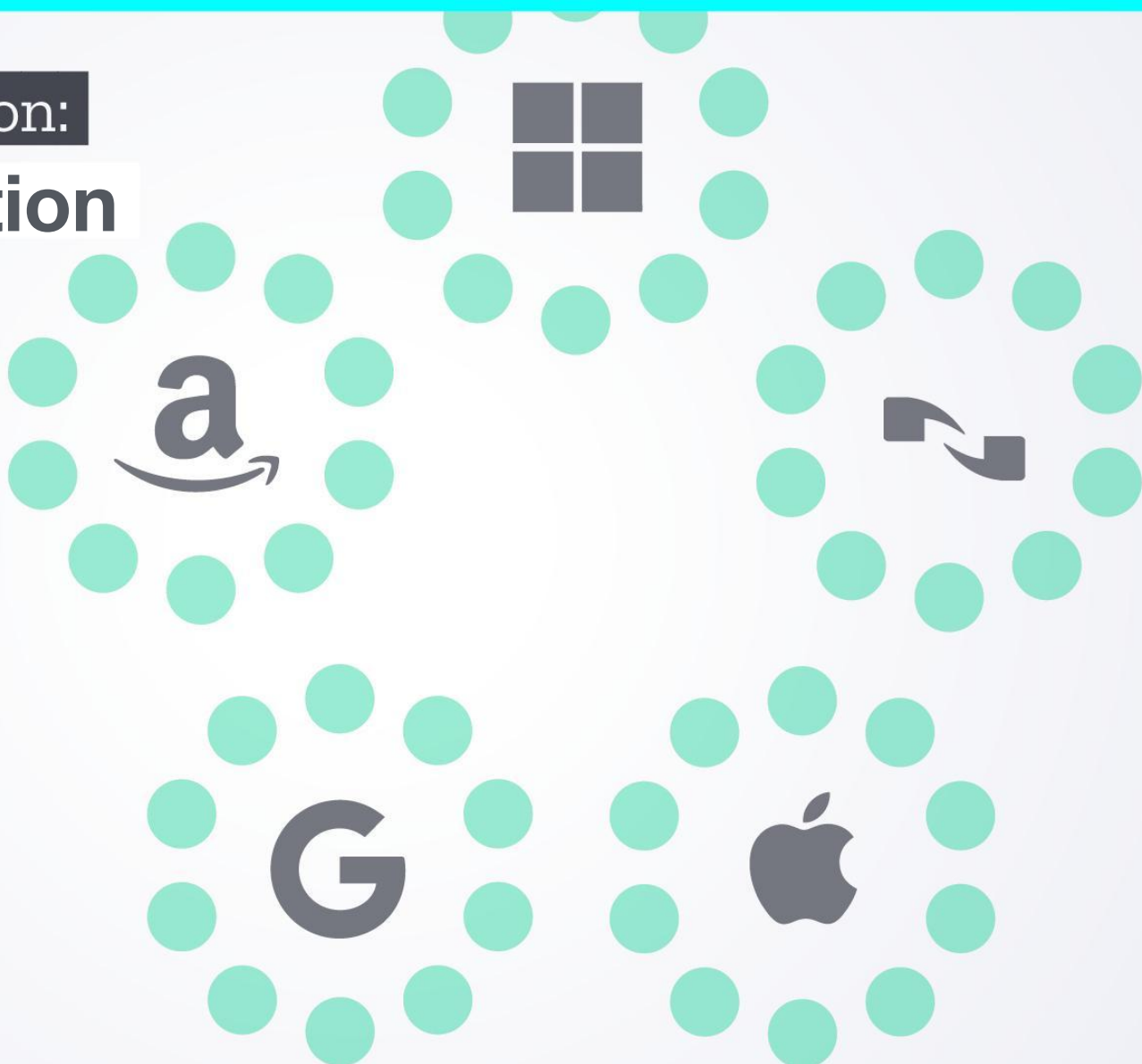
Speech Recognition:

Voice Assistants (US)



Speech Recognition:

Proliferation



Speech Recognition

Ingredients



Project **Common Voice**

Project **Deep Speech**

III

Common Voice



Common Voice

voice.mozilla.org

Progetto di Mozilla incentrato sul raccoglimento di registrazioni vocali, per poter creare un riconoscitore vocale basato su un dataset e su un modello di pubblico dominio per tutte le lingue che partecipano al progetto.

Il modello di deep learning, basato su un paper di Baidu, è implementato dal progetto DeepSpeech di Mozilla.

<https://github.com/mozilla/voice-web>

moz://a

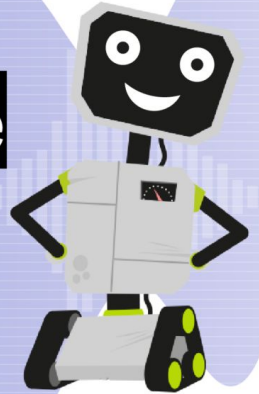
emerging

technologies



Common Voice

moz://a



Perché è nato il progetto?

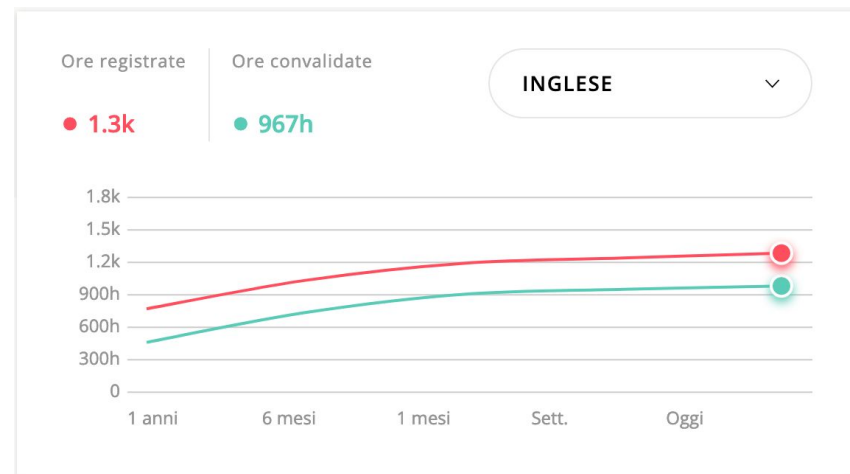
- ❑ Contrastare la centralizzazione dei motori Speech-To-Text proprietari
- ❑ Contrastarli perché funzionano solo online con problemi legati alla privacy degli utenti
- ❑ Rendere il riconoscimento vocale disponibile per tutte le lingue in maniera open source
- ❑ Anche per le lingue con meno interesse commerciale
- ❑ Non avere limitazioni sull'utilizzo di questi servizi



Stato del progetto per la lingua inglese

- ❑ 39577 partecipanti
- ❑ 30 GB di dataset
- ❑ WER versione 0.2.0: **11%**

Nell'ultimo anno le ore convalidate sono passate da 496 alle attuali 968

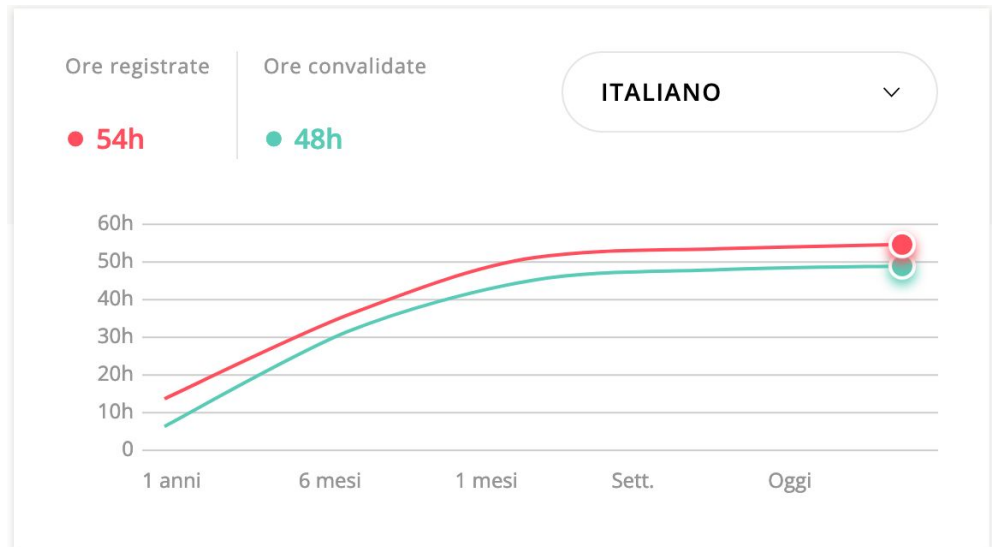
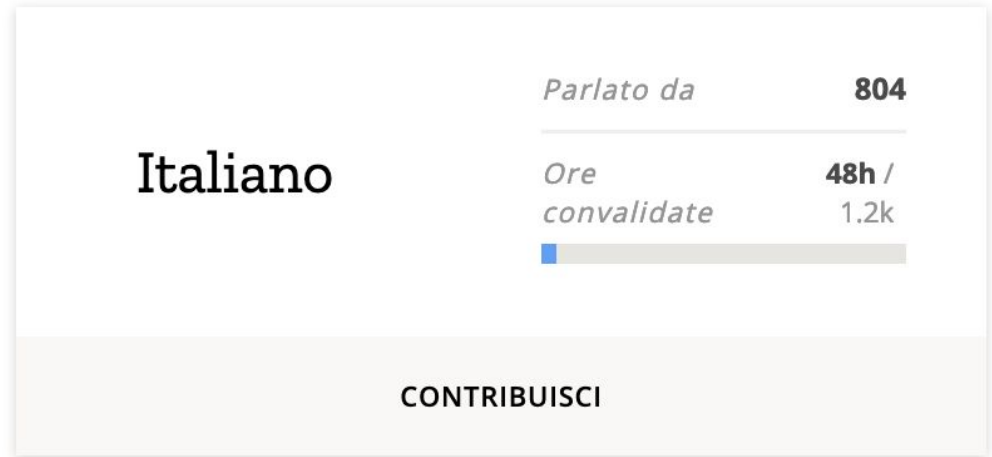


Stato del progetto per la lingua italiana

- ❑ Oltre 11000+ frasi da diverse fonti
- ❑ Avviato il 3 luglio 2018
- ❑ Promozione:
 - ❑ Social network
 - ❑ Linux Day
 - ❑ Italian Hacker Camp
 - ❑ Meetup in tutta Italia

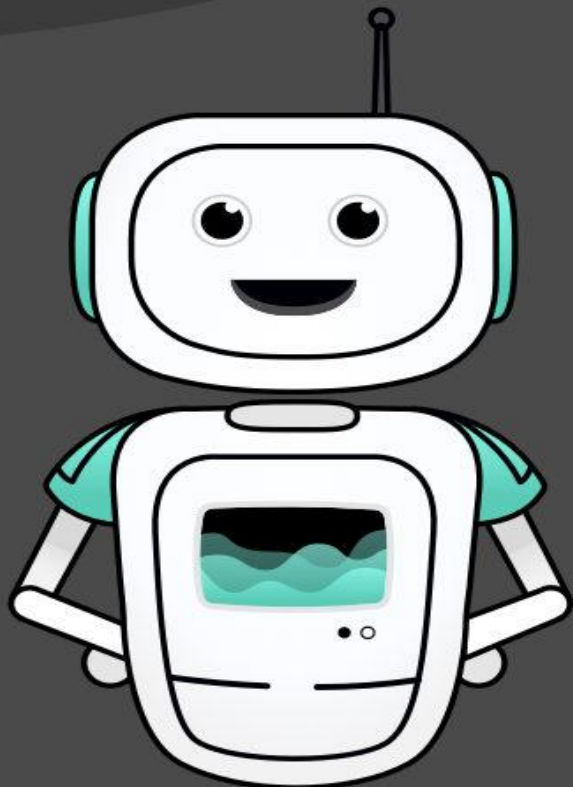
Nell'ultimo anno da 6,5 ore a 48!

I partecipanti da 181 a 804!



Common Voice

moz://a



Validated Hours

1965



Recorded Hours

2454



Languages

29



Stiamo

realizzando un dataset di voci multilingua e open source che chiunque può utilizzare per migliorare applicazioni munite riconoscimento vocale.

Crediamo che mettere in circolazione dataset di voci pubblici di grandi dimensioni possa promuovere innovazione e sana concorrenza commerciale tra le tecnologie di apprendimento automatico basate sul riconoscimento vocale.

Il dataset multilingue di Common Voice è già il più grande del suo genere disponibile pubblicamente, ma non è l'unico.

Considera questa pagina come un punto di riferimento per dataset open source e, man mano che Common Voice continua a crescere, una pagina per i nostri aggiornamenti.

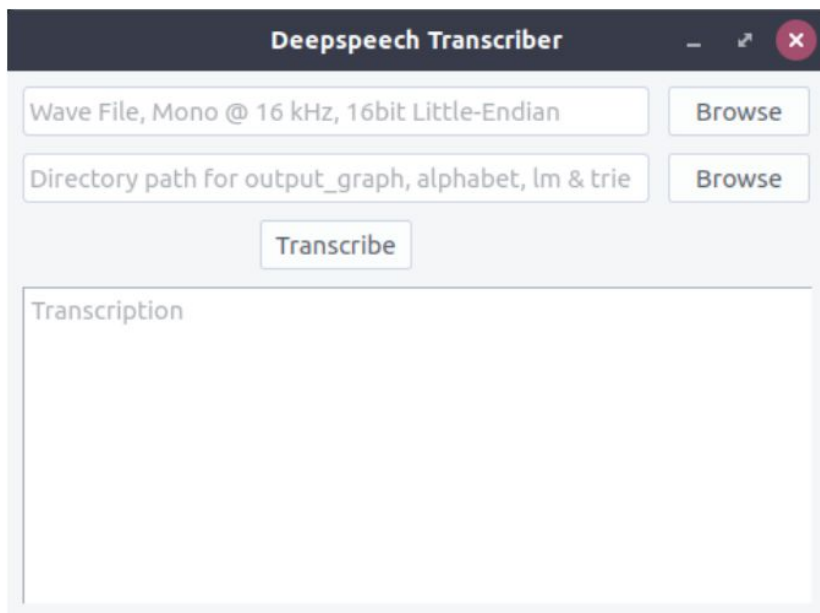


Lingua
Italiano

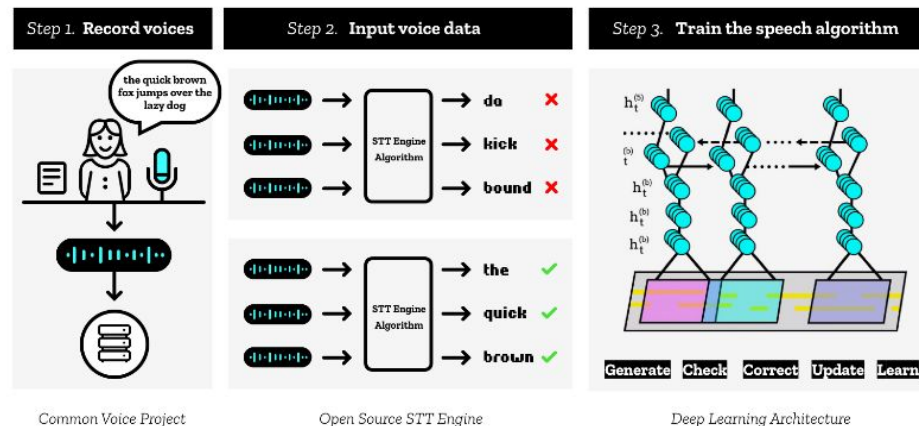
DIMENSIONE	1 GB
VERSIONE	it_40h_2019-06-12
TOTALE ORE CONVALIDATE	36
TOTALE ORE COMPLESSIVE	40
LICENZA	CC-0
NUMERO DI VOCI	602
FORMATO AUDIO	MP3
DIVISIONI	Accento Età 32% 19 - 29, 19% 50 - 59, 11% 30 - 39, 10% 40 - 49, 6% < 19 Sesso 62% Maschio, 18% Femmina

DeepSpeech: modello

- già scaricabile, versione 0.5.1 (0.6 soon)
- disponibile il pre-trained in inglese
- Affidabilità maggiore del 94%
(Novembre 2017)



How a Speech Application Learns



Prima versione modello DeepSpeech in Italiano

- ❑ Immagine Docker per generare modello
- ❑ Training con dataset Common Voice e Open-Corpora (~170 ore)
- ❑ Versione 0.1a modello pubblicata: pronta per essere testata!!

<https://github.com/MozillaItalia/DeepSpeech-Italian-Model>

Pre-release

model-0.1a

2a550f4

Verified

Trained models 0.1a

 mone27 released this 9 days ago · 2 commits to master since this release

First version of Italian model, trained with:

- ~40 hours for Common Voice IT dataset
- ~ 127 hours of m-ailabs Italian dataset
total: ~170h

model hyper-parameters:

- batch_size=68
- n_hidden=2048
- epochs=30
- learning_rate=0.00025
- dropout=0.15
- lm_alpha=0.75
- lm_beta=1.85
- early_stop=0

IV

Contribuisci!



Come contribuire a Common Voice?

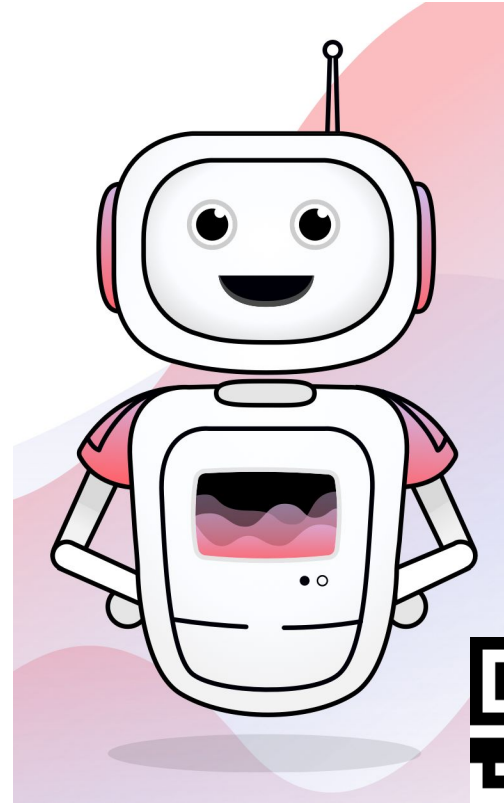
Registra la tua voce

- Leggi la frase che compare
- Scandisci tutte le parole, anche le finali!
- Leggi anche i numeri, per esteso
- **Non preoccuparti del tuo accento!**

Revisiona le registrazioni

Accertati che le registrazioni siano corrette:

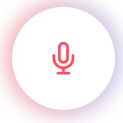
- Frasi lette per intero e correttamente
- Deve essere comprensibile/udibile, quindi **assenza**, o bassi, di rumori di sottofondo!



1. Vai su: voice.mozilla.org/it

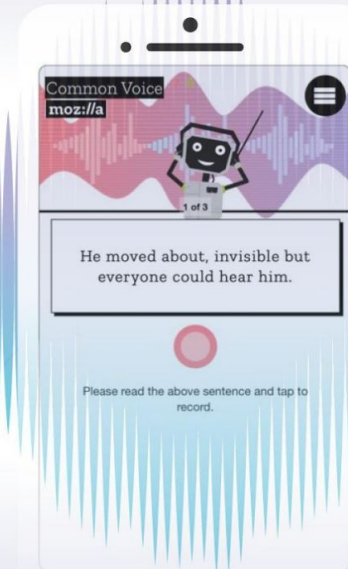
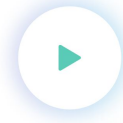
Parla

Dona la tua voce



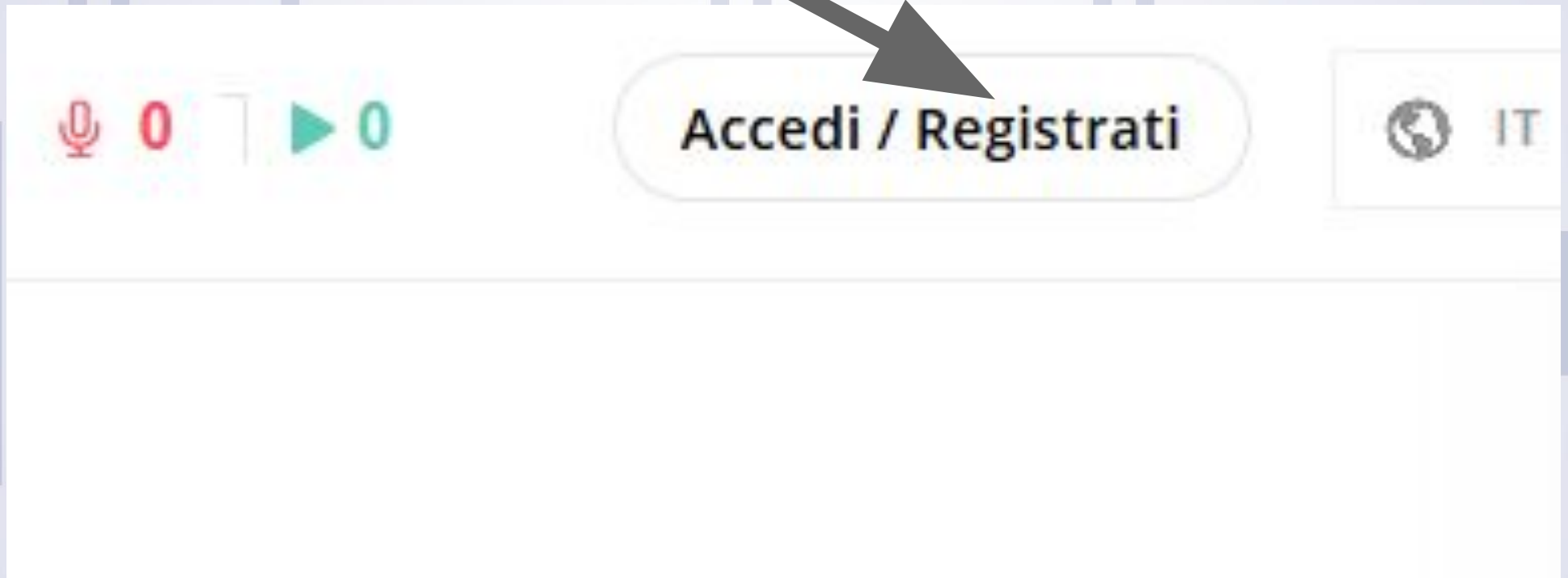
Ascolta

Aiutaci a convalidare le
registrazioni



Sia da PC che da smartphone

2. Crea un account



<https://voice.mozilla.org/it>

3. Dona la tua voce!

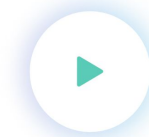
Parla

Dona la tua voce



Ascolta

Aiutaci a convalidare le
registrazioni



Obiettivo: 15 Minuti (225 registrazioni) a persona

4. Convalida le voci!

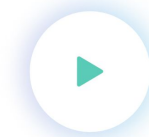
Parla

Dona la tua voce



Ascolta

Aiutaci a convalidare le
registrazioni



Obiettivo: 450+ revisioni a persona

5. Controlla i tuoi progressi



5



1

[Log In / Sign Up](#)

 **225**  **450**

[Log In / Sign Up](#)

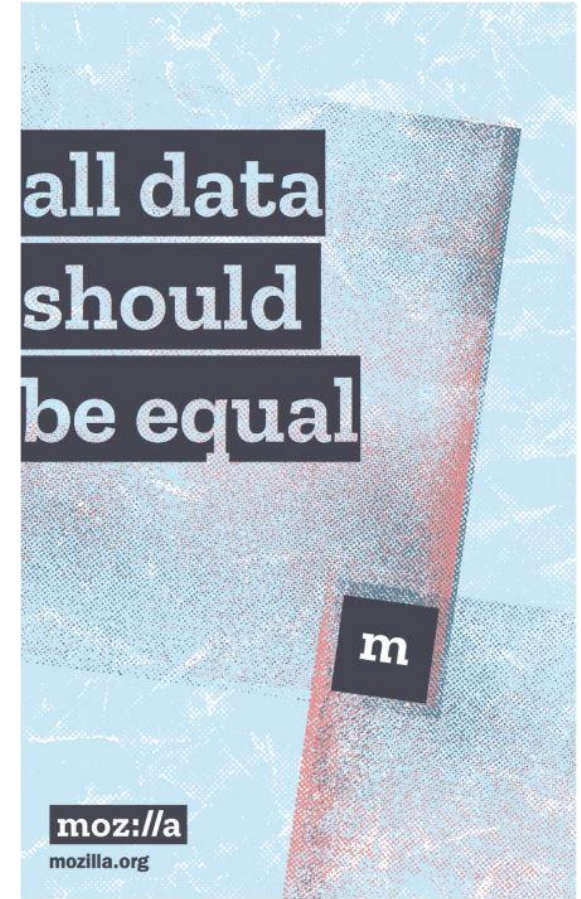
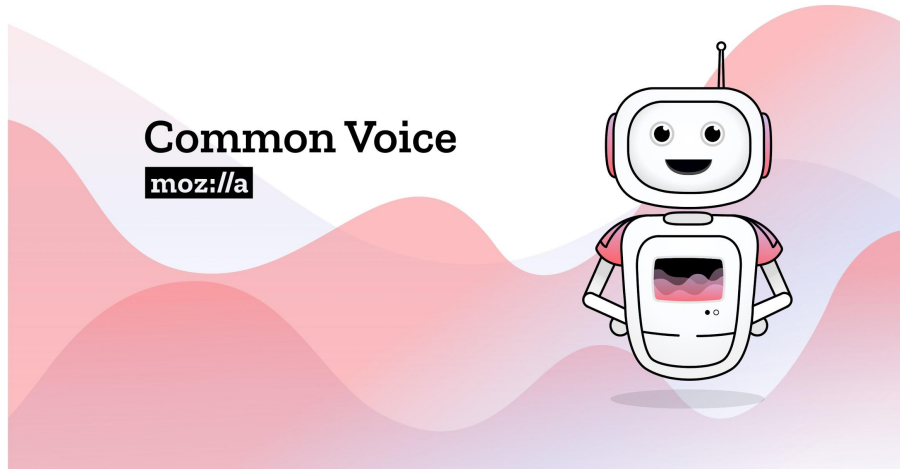
Festeggia!



WE
NEED
YOU

Obiettivi per il futuro di Common Voice

- aggiungere il supporto per altre lingue
- migliorare le performance in caso di rumore di fondo
- proposta per API web speech
- continuare a migliorare l'interfaccia
- aggiungere più funzionalità per tracciare lo stato di raccolta dei dati vocali

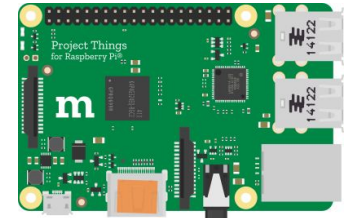


DeepSpeech/Common Voice: utilizzi

- Mozilla IoT:
assistente sperimentale per il
Web of Things Gateway
- Mycroft AI:
uso del dataset di Common Voice
per lo sviluppo di un assistente
vocale completamente Open Source

moz://a

IoT



MYCROFT AI

Grazie!

luisa.ravelli@lugbs.linux.it

Per restare aggiornato sulle novità di Mozilla unisciti ai canali e ai gruppi ufficiali su Telegram: @mozItabot

Sito: [mozillaitalia.org](https://www.mozillaitalia.org)

moz://a