

UNIVERSITÀ DEGLI STUDI DI PARMA

Corso di Laurea in Ingegneria dei Sistemi Informativi

Tesi di Laurea di primo livello

Realizzazione di un assistente vocale per il triage ospedaliero

Triage semi-automatizzato per la sicurezza e la tempestività



Relatore

prof.ssa Mordonini Monica

Correlatori:

prof. Tomaiuolo Michele

prof. Angiani Giulio

Laureando

Simone Montali

matricola 288144

Probabilmente 2020

Dedica toccante.

Citatemi dicendo che sono stato citato male.
Groucho Marx

Ringraziamenti

Grazie

Sommario

Abstract della tesi

Indice

Elenco delle figure	VII
1 Introduzione	1
2 Stato dell'arte	2
2.1 Machine Learning	2
2.1.1 Reti neurali	3
2.2 Assistenti vocali	4
2.2.1 Assistenti vocali in ambito medico	5
2.3 Triage	7
2.3.1 Come il COVID19 ha modificato le procedure di triage	7
3 Mycroft	9
3.1 Open source	9
3.1.1 Cosa si intende con open source?	9
3.1.2 Perché open source?	10
3.2 Introduzione a Mycroft	10
3.2.1 Stack di funzionamento	10
A An appendix	12
Riferimenti bibliografici	14

Elenco delle figure

2.1	Modello lineare	3
2.2	Modello a 2 strati, ancora lineare	3
2.3	Rete neurale	4
2.4	Percentuali di utenti che hanno utilizzato/vorrebbero utilizzare assistenti vocali in ambito medico	5
2.5	Interesse dei consumatori negli assistenti vocali in ambito medico, per fasce di età	6
2.6	Schema del protocollo CESIRA	7

Capitolo 1

Introduzione

In una società sempre più digitalizzata e dinamica, spesso viene a crearsi un netto distacco tra i settori capaci di **evolvere assieme alla tecnologia**, e quelli che, per un motivo o per l'altro, rimangono ancorati a procedure e metodologie tradizionali. Il settore medico, in continuo rinnovamento sul lato scientifico, è, soprattutto in Italia, affidato ad **infrastrutture informatiche progettate tempo fa**. Questo, soprattutto per motivi di stabilità e affidabilità: gli errori, qui, non sono ammissibili.

Per questo motivo, spesso non si notano le evidenti possibilità di miglioramenti che la ricerca informatica potrebbe apportare agli ospedali, agli ambulatori, agli studi. L'obiettivo di questa tesi è mettere luce su una delle possibili modalità con cui l'informatica potrebbe, in un futuro prossimo, migliorare la praticità ma soprattutto la sicurezza degli ambienti ospedalieri.

La procedura di triage, ossia il processo di selezione dei pazienti richiedenti cure, è oggi affidata totalmente ad infermieri. Questa scelta è dovuta, oltre ad un evidente bisogno di poter osservare il paziente, alla necessità dell'**immediatezza di un contatto verbale** con il personale sanitario.

Perciò, la sfida nella realizzazione della tesi è soprattutto legata al rendere il **più umana possibile** l'interazione con un bot automatizzato. Il bot deve quindi accogliere il paziente, comprenderne le problematiche ed i sintomi, e farlo sentire protetto. Non si escluderà del tutto un apporto umano: gli infermieri sono addestrati per poter osservare e comprendere il richiedente cura, e l'apporto dell'osservazione diretta del paziente è ancora troppo importante per escluderla. È senz'altro possibile, però, affidare la prima parte di **profilazione dell'utente** a procedure automatizzate.

Capitolo 2

Stato dell'arte

Illustriamo ora lo stato dell'arte delle varie tecniche e tecnologie che coinvolgono il bot.

2.1 Machine Learning

Il machine learning (in seguito, ML) è una disciplina che si occupa di rispondere, sostanzialmente, a due domande: *Come si può costruire un sistema informatico che migliora con l'esperienza?* e *Quali sono le leggi teoriche fondamentali che governano ogni sistema di apprendimento, sia esso implementato nei computer, negli umani, nelle organizzazioni?*. Il ML copre una serie di compiti diversi, dalla classificazione di email spam, al riconoscimento facciale, al controllo di robot [4]. Ogni problema di ML può essere riassunto come il miglioramento di una determinata performance (data, ad esempio, dall'errore sulle predizioni) durante l'esecuzione di un'operazione. Seguendo l'esempio delle email spam, un algoritmo di ML potrebbe apprendere, tramite un *training set* di email già etichettate come spam/non-spam, a categorizzarne altre. Portando il problema in termini matematici, vogliamo, dato un grafico in cui i punti rossi sono email spam, e quelli blu no, trovare una funzione che ci permetta di evidenziare le email da eliminare. Ciò che faremo sarà quindi definire un insieme di **features**, consistenti, ad esempio, nel contenuto testuale della mail (vettorizzato in termini numerici), l'orario di arrivo, la lunghezza in parole... Oltre a ciò, definiremo una **label**, ossia la caratteristica che vogliamo predire: sarà, per esempio, 1 se l'email è spam, 0 altrimenti. Ora, il problema consiste in una semplice regressione lineare: date le features x_0, x_1, x_2 e la label y , vogliamo trovare dei pesi w_0, w_1, w_2 tali che

$$w_0x_0 + w_1x_1 + w_2x_2$$

approssimi al meglio y . Per fare ciò, procediamo per iterazioni: partiremo con dei pesi arbitrari, calcolando y per ogni esempio (già etichettato) fornito, e ne calcoleremo l'errore rispetto alla y reale. Potremo quindi calcolare il gradiente (ossia la somma delle derivate parziali sui rispettivi pesi) della funzione che lega l'errore ai pesi w_i , ottenendone un'informazione fondamentale: indicherà infatti la direzione nella quale la funzione errore è decrescente. Ci potremo quindi spostare, di una quantità pari al *learning rate* (uno degli iperparametri definiti arbitrariamente), nella direzione di diminuzione dell'errore. Scegliere un corretto *learning rate* è una parte fondamentale dello sviluppo dell'algoritmo di ML:

se è troppo piccolo, la soluzione richiederà troppo tempo, se troppo grande, la soluzione non sarà abbastanza precisa. Dopo un sufficiente numero di esecuzioni dell'algoritmo, il nostro grafico avrà ora questo aspetto: e saremo quindi in grado di determinare, con una determinata *accuracy*, l'appartenenza di un esempio ad una classe o all'altra.

2.1.1 Reti neurali

Per illustrare il concetto di rete neurale, produciamo anzitutto un grafico del modello lineare appena definito.

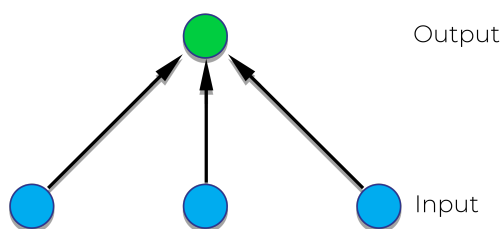


Figura 2.1: Modello lineare

Ogni cerchio blu rappresenta una feature, ed il cerchio verde rappresenta la somma pesata degli input. Vorremmo che questo modello fosse in grado di approssimare anche problemi non-lineari. Potremmo quindi decidere di inserire un nuovo strato di valori intermedi, ottenendo un modello così composto:

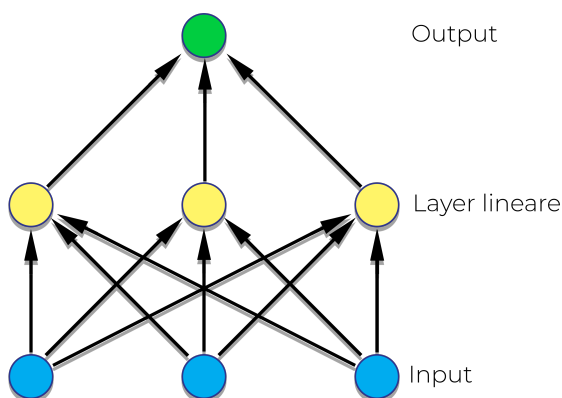


Figura 2.2: Modello a 2 strati, ancora lineare

c'è, però, un problema: essendo questo ancora un modello lineare, è possibile semplificarlo, ottenendo il primo grafico. È utile, allora, inserire una non-linearità all'interno del modello, che chiameremo *funzione di attivazione*: ora avremo, oltre ai layer lineari, una funzione tra di loro non lineare. Grazie a essa, l'aggiunta di nuovi strati è diventata significativa: non sarebbe più possibile semplificare il sistema ad uno lineare.

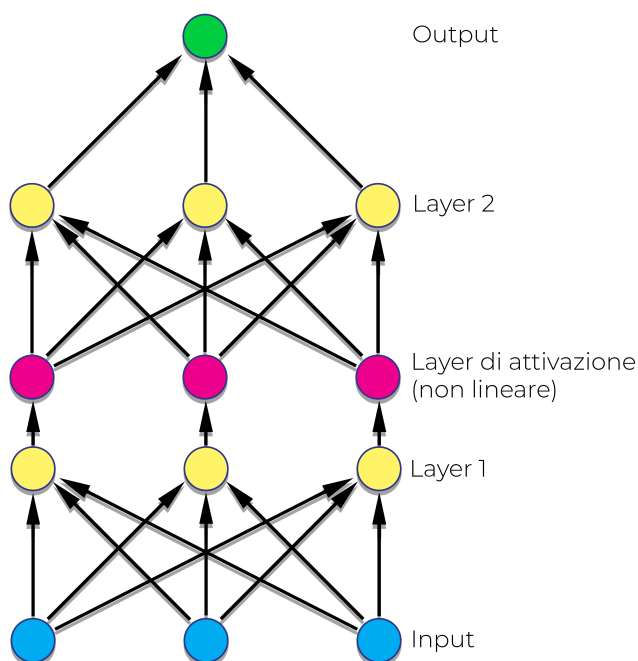


Figura 2.3: Rete neurale

Tra le funzioni di attivazione, citiamo la funzione sigmoidea $F(x) = \frac{1}{1+e^{-x}}$, o la *Rectified Linear Unit* (ReLU) $F(x) = \max(0, x)$. Non esiste una funzione di attivazione più corretta delle altre: ogni problema avrà una configurazione dei layer e delle loro attivazioni diversa. Trovare la configurazione corretta della rete neurale è la sfida più grande nel lavoro di *data analysis*.

Avremo ora quindi un modello capace di ottenere degli output date le input features. Ma come possiamo addestrarlo per ottenere dei risultati validi? Una soluzione ricorrente è quella della **retropropagazione dell'errore**, o *backpropagation*. Durante questo processo, la rete neurale calcola il gradiente dell'errore (ottenuto da una funzione di *loss*) sull'output, rispettivamente ai vari pesi degli strati. Questo permette alla rete neurale di individuare quali nodi hanno più effetto sulla generazione dell'errore, e di conseguenza, regolarne i pesi per ridurre l'errore. Fatto ciò sullo strato più vicino all'output, si tratta solo di ripetere il procedimento "*a salire*" per ogni layer della rete neurale.

2.2 Assistenti vocali

Gli assistenti vocali sono software in grado di interpretare dialoghi umani e rispondere attraverso voci sintetizzate. Alcuni esempi famosi sono Apple Siri, Amazon Alexa, Google Assistant. Gli utenti possono effettuare domande, controllare dispositivi domotici, e svolgere un numero di operazioni in continua espansione [3]. Il software è costantemente in attesa di una *wake word*; una volta riconosciuta questa parola, si mette in ascolto di comandi. La richiesta viene quindi tradotta in testo tramite un motore di *Speech To Text* ed interpretata dal software. Questa è senz'altro la parte più complessa: il compito di interpretazione di

linguaggio naturale è reso complesso dalle svariate sfaccettature che ogni diversa lingua presenta. Interpretata la richiesta, l'assistente procede con le operazioni collegate, e prepara una risposta testuale da dare all'utente. Questa risposta verrà poi sintetizzata in audio da un motore di *Text To Speech*. I moderni dispositivi di assistenza vocale come il Google Home o l'Amazon Echo si appoggiano a server in cloud per l'interpretazione delle richieste: questo permette di ridurne notevolmente le dimensioni ed i requisiti tecnici. Lo sviluppo, negli ultimi anni, di nuove tecniche di ML, unite ad un grande incremento delle potenze di calcolo e alla disponibilità di vastissimi *dataset linguistici*, hanno permesso grandi miglioramenti nella tecnologia, ma soprattutto l'applicazione ad ambiti, fino a qualche anno fa, impensabili.

2.2.1 Assistenti vocali in ambito medico

Negli anni sono emerse alcune possibili applicazioni delle tecnologie per assistenti vocali in ambito medico. La startup **Saykara** ha creato un software che permette di ascoltare le conversazioni medico-paziente, allo scopo di creare schede cliniche accurate senza dover impegnare il medico in lunghe procedure di *data entry*. Un report [5] di Voicebot, organizzazione per la diffusione degli assistenti vocali, afferma che il 52% degli intervistati sarebbe intenzionato ad utilizzare assistenti vocali in ambito medico nel futuro. Solo un 7.5% (1 su 13) l'ha già fatto.

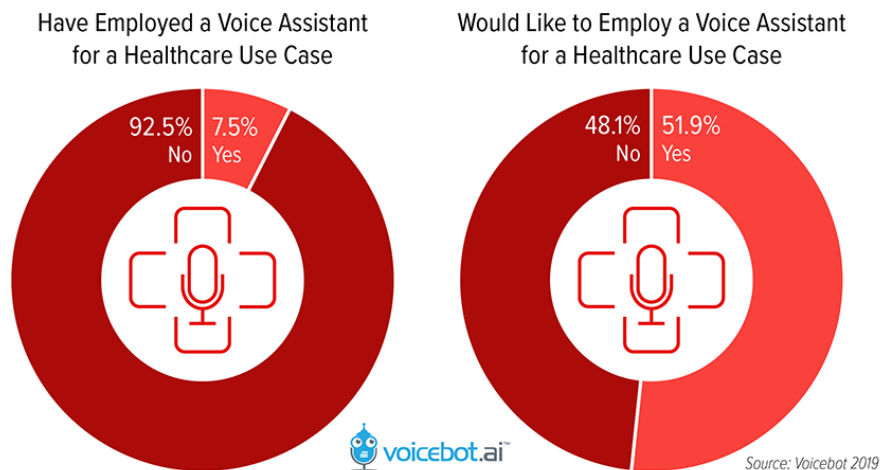


Figura 2.4: Percentuali di utenti che hanno utilizzato/vorrebbero utilizzare assistenti vocali in ambito medico

È interessante notare come anche le fasce di età più avanzate siano, seppur con percentuali più basse, interessate a provare simili applicazioni della tecnologia.

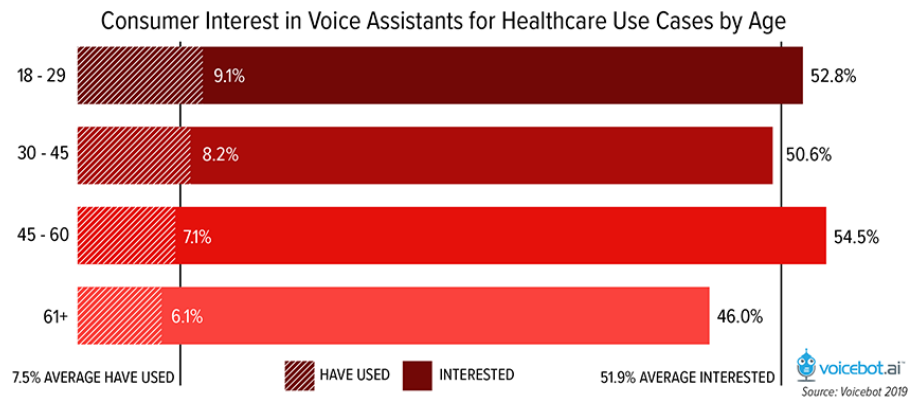


Figura 2.5: Interesse dei consumatori negli assistenti vocali in ambito medico, per fasce di età

Un altro esempio interessante di applicazione della tecnologia in ambito medico è *Dragon Medical Virtual Assistant* rilasciato da **Nuance**, una suite di strumenti vocali capaci di creare cartelle cliniche, eseguire diagnostica sul paziente, fornire risultati in diretta di ricerche mediche, supportare le operazioni di radiologia.

2.3 Triage

Il **triage** è, per definizione, il sistema utilizzato per selezionare i soggetti coinvolti in infortuni, gravi o leggeri che siano, secondo classi di emergenza, in base alla gravità delle lesioni riportate o del loro quadro clinico. Il termine nacque in ambiti di guerra (se ne cominciò a parlare durante le *guerre napoleoniche*), durante i quali si classificavano i soldati feriti in base alla probabilità di sopravvivenza. Il procedimento venne poi, negli anni, iniziato ad utilizzare in ambito ospedaliero, con l'applicazione di metodi più scientifici come il modello *START*, basato su un algoritmo che prevede la verifica di sintomi visibili, frequenza respiratoria, pulsazioni, risposta ai comandi. Sebbene inizialmente le procedure di triage fossero sfruttate solo in ambito di maxiemergenze, oggi tutti i pronto soccorso italiani ne fanno uso in qualche modo. Non esiste però uno standard di codici utilizzati in ogni ospedale: le normative cambiano tra un pronto soccorso e l'altro. Distinguiamo però, generalmente, 4 tipi di codice:

- Codice **rosso**: paziente critico con cedimento di una o più funzioni vitali. *Accesso immediato.*
- Codice **giallo**: potenziale rischio di vita o compromissione di una o più funzioni vitali. *Accesso entro 15 minuti.*
- Codice **verde**: valutazione medica differibile nel tempo senza alcun rischio evolutivo. *Accesso entro 60-90 minuti.*
- Codice **bianco**: paziente senza emergenza, con problematiche risolvibili a domicilio o dal medico di famiglia.

Il triage infermieristico produce quindi una **scheda di triage**, che ha valore giuridico, contenente anagrafica, data/ora di accettazione, motivo, codice colore. Durante le emergenze è comune utilizzare il protocollo **CESIRA** (Coscienza, Emorragia, Shock, Insufficienza respiratoria, Rotture, Altro).

2.3.1 Come il COVID19 ha modificato le procedure di triage

La pandemia mondiale di COVID19 ha modificato i processi di triage in ospedale: lo scopo è limitare al massimo le possibilità di contagio tra i pazienti, rendendo il pronto soccorso più sicuro tanto per i pazienti, quanto per il personale sanitario. Alcuni ospedali in Inghilterra hanno addirittura iniziato a praticare triage remoto: il paziente comunica le sintomatologie online

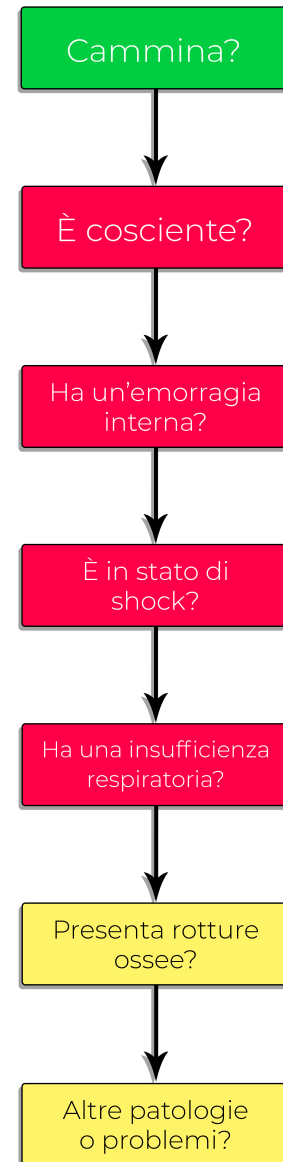


Figura 2.6: Schema del protocollo CESIRA

e, solo se necessario, può recarsi in ospedale per la visita medica [2]. In Italia, ogni ospedale ha provveduto a definire delle procedure di **pre-triage** atte ad evitare il contatto tra pazienti affetti da COVID19 e non. Generalmente, si tratta di percorsi obbligati in cui i pazienti con sintomi compatibili con COVID19 (febbre, tosse, insufficienza respiratoria) vengono separati in un'*area di decontaminazione* e visitati senza accedere agli ambienti del pronto soccorso comune. Nell'ospedale Maggiore di Parma sono state previste, durante i momenti di picco dell'epidemia, due tende di pre-triage esterne all'ospedale stesso.

Capitolo 3

Mycroft

There's an entire community of developers looking to access this technology, but so far, it's been the purview of a few large companies. The technology is walled-off, proprietary, and secretive

Joshua Montgomery
CEO di Mycroft AI, Inc.

Nel panorama degli assistenti vocali disponibili al pubblico, le grandi aziende catturano la maggior parte delle attenzioni dei consumatori. I nomi di Amazon Alexa, Google Home, Microsoft Cortana sono familiari a tanti. Meno familiare, è, però, lo scenario degli stessi strumenti in ambito di assistenti vocali. Piattaforme come *Jarvis*, *Linto*, *Open Assistant* e **Mycroft** sono sconosciute ai più. Quest'ultimo ha, negli ultimi anni, catturato l'attenzione di molti esperti di settore per la sua stabilità, espansibilità, popolarità.

3.1 Open source

3.1.1 Cosa si intende con open source?

Il termine **open source** si riferisce a qualcosa di modificabile e condivisibile dalle persone, avente quindi un **design pubblicamente accessibile**. Il termine nacque nell'ambito dello sviluppo software, ma oggi rappresenta piuttosto un'etica. Il software open source è software il cui codice sorgente è ispezionabile, modificabile, e migliorabile da chiunque voglia farlo. Con *codice sorgente* intendiamo il codice che definisce il comportamento del programma, ossia il prodotto del lavoro di un programmatore.

Molti dei software che oggi utilizziamo sono invece **closed source** (o *proprietario*), ossia software il cui codice sorgente non è accessibile agli utenti, e sul quale l'organizzazione che crea il software ha *pieni poteri*. Alcuni esempi di software proprietario possono essere Windows, Photoshop, Safari. Solitamente, durante l'installazione di uno di questi software,

l'utente accetta dei termini e condizioni che lo vincolano a non fare nulla di non autorizzato dall'azienda creatrice del software.

3.1.2 Perché open source?

Le motivazioni per preferire software open source rispetto a quelli proprietari sono svariate. Sebbene i software proprietari spesso abbiano funzionalità più avanzate e meglio sviluppate, hanno gravi mancanze in progetti come questo. Il software open source, a scapito della possibilità di minori funzionalità (nonostante ciò non sia sempre vero, basti pensare alla potenza del kernel Linux), dà ad un utente esperto tante possibilità:

- Possibilità di **controllo** sul software: la modifica e il miglioramento del software è permessa e ben vista. La comunità open source lavora con coesione per produrre software sempre migliore.
- Possibilità di conoscere e studiare i **meccanismi interni**: l'utilizzo di software per fini scientifici richiede la piena conoscenza di come un software funziona. Inoltre, la disponibilità del codice a tutti permette agli studenti e agli interessati di scoprire le logiche che ne permettono il corretto funzionamento.
- Grandi garanzie di **sicurezza**: se tutti possono accedere al codice, la presenza di bug di sicurezza e problematiche di privacy viene rilevata molto più velocemente di come accade nel software proprietario.
- Certezze sulla **privacy**: la società in cui viviamo è sempre più basata sulla vendita di dati personali acquisiti tramite software. L'open source, spesso e volentieri, si oppone a tale tendenza: essendo software della comunità e non di un'azienda che punta all'ottenimento di capitale, nessuno ha interesse nel *monetizzare gli utenti*.
- Supporto della **comunità**: lo sviluppo di software open source è molto supportato dalla comunità informatica. In caso di problematiche o necessità, ci sarà sempre un utente più esperto disposto ad aiutare.
- Correttezza **etica**: software come quello oggetto di questa tesi hanno l'obiettivo di salvare vite. È giusto che tutti possano accedervi, migliorarlo e conoscerlo. La ricerca scientifica dovrebbe essere il più possibile **libera**.

3.2 Introduzione a Mycroft

Mycroft nacque tramite crowdfunding nel 2015 e catturò subito molto interesse da parte della comunità informatica. L'obiettivo era chiaro: opporsi alla sola presenza di software proprietario in ambito open source. Ricevette presto l'appoggio di molte organizzazioni come la Canonical (promotrice di Ubuntu) e la Mozilla Foundation. La repository su Github di Mycroft conta oggi *4300 stars* in continua crescita.

3.2.1 Stack di funzionamento

Come spiegato nel capitolo 2.2, un assistente vocale si compone principalmente di:

Rilevamento della wake word L'assistente ha bisogno di rilevare una parola per attivarsi. Gli esempi più famosi sono *Hey, Google* o *Alexa*. Mycroft permette di personalizzare la propria wake word, che all'installazione è semplicemente *Hey, Mycroft*. Il progetto utilizza **Precise**, un wake word listener basato su reti neurali addestrate su esempi sonori. Questo componente, totalmente open source, è basato su *pattern sonori*, non sul testo. Questo ne riduce la dipendenza da accenti e linguaggi diversi. Precise offre la possibilità di addestrare il listener su un proprio dataset.

Speech To Text Mycroft sfrutta il motore STT di Google per effettuare questo processo. Per aggiungere uno strato di privacy a queste richieste, esse vengono fatte passare attraverso i server Mycroft: Google non può ricollegare una richiesta all'utente che l'ha fatta. In questo modo, tutte le query effettuate a Google contengono solamente l'audio e hanno lo stesso mittente. In futuro sarà possibile sfruttare il dataset di **Mozilla Deepspeech**, che non è, ad oggi, ancora sufficientemente affidabile.

Interpretazione degli intent Nell'ambito della speech recognition, un intent è l'operazione che l'utente *intende* svolgere. Un utente può richiedere un'operazione in modi molto diversi. Lo scopo dell'intent parser è proprio quello di riuscire a superare questo scoglio, estraendo dalla richiesta scritta gli elementi chiave. Per esempio, se avessimo una richiesta del tipo "Hey Mycroft, domani pioverà a Parma?", l'intent parser dovrebbe capire:

- L'utente desidera conoscere il **tempo atmosferico**
- Il tempo atmosferico deve essere cercato per **Parma**
- La data di interesse è **domani**

Mycroft rende disponibili due software per la rilevazione di intenti: Padatious e Adapt. Mentre il secondo è basato sul riconoscimento di parole chiave, ed è quindi molto suscettibile alle richieste, il primo è composto da una rete neurale addestrata su frasi intere. Nel corso del progetto sfrutteremo **Padatious**.

Text To Speech Il componente di Text To Speech ha il compito di generare una traccia audio partendo dal testo di risposta codificato nella skill. Mycroft rende disponibile il progetto **Mimic**, basato sul software FLITE dell'università Carnegie Mellon. Purtroppo quest'ultimo non è disponibile in italiano, quindi nel corso del progetto utilizzeremo il motore TTS di Google, ottimizzato per l'italiano. Come sempre, le richieste passano attraverso i server Mycroft per ottenere un buon livello di privacy.

Appendice A

An appendix

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text

will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Bibliografia

- [1] Geoffrey I. Webb (Editor) Claude Sammut (Editor). *Encyclopedia of Machine Learning and Data Mining*. 2^a ed. Springer, 2017. ISBN: 148997685X, 978-1489976857.
- [2] Piers Ford. «Digital triage changes the game for UK GP practices in the age of COVID-19». In: (mag. 2020). URL: <https://www.mobihealthnews.com/news/europe/digital-triage-changes-game-uk-gp-practices-age-covid-19>.
- [3] Matthew Hoy. «Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants». In: *Medical Reference Services Quarterly* 37 (gen. 2018), pp. 81–88. DOI: [10.1080/02763869.2018.1404391](https://doi.org/10.1080/02763869.2018.1404391).
- [4] T.M. Mitchell. *Machine Learning*. McGraw-Hill International Editions. McGraw-Hill, 1997, pp. 81–88. ISBN: 9780071154673. URL: <https://books.google.it/books?id=EoYBngEACAAJ>.
- [5] Voicebot - Orbita. «Voice Assistant Consumer Adoption in Healthcare». In: (ott. 2019). URL: https://voicebot.ai/wp-content/uploads/2019/10/voice_assistant_consumer_adoption_in_healthcare_report_voicebot.pdf.