

Natural Language Processing e Computer Vision per l'accessibilità dei video di lezioni con formule

Tiziana Armano¹, Gabriele Baratto³ Cristian Bernareggi¹, Sara Bertoldo⁴, Erika Brunetto¹, Anna Capietto¹, Manuela Caramagna³, Sandro Coriasco¹, Nazareno Defrancesco⁴, Mattia Ducci¹, Maria Luisa Gabrielli⁴, Mariafrancesca Guadalupi⁴, Alessandro Mazzei², Angelo Saccà³, Francesco Tarasconi⁴

¹Dipartimento di Matematica – Università di Torino, ²Dipartimento di Informatica – Università di Torino, ³Direzione SIPE - Università di Torino, ⁴Maize srl

Abstract. Il contributo presenta il progetto VoiceMath (iniziato nel 2021) del Laboratorio Polin per la ricerca e la sperimentazione di nuove tecnologie assistive per le STEM, in collaborazione con Maize srl e la Direzione SIPE dell'Università di Torino. VoiceMath è un'applicazione web sviluppata per rispondere alle esigenze di accessibilità di video di lezioni universitarie di contenuto scientifico. Punti di forza del progetto sono: l'efficace trasformazione di un prodotto della ricerca in servizio per l'ateneo, l'impiego di tecnologie innovative per lo sviluppo di uno strumento complesso, l'evoluzione delle funzionalità in relazione ai nuovi sviluppi delle tecnologie utilizzate.

Keywords. LLM, NLP, computer vision, accessibilità, formule, speech-to-text

Introduzione

Il periodo pandemico ha accelerato il processo di innovazione digitale nella didattica aumentando l'utilizzo di strumenti e risorse digitali ormai di uso comune. Tuttavia, l'utilizzo di strumenti e risorse digitali ha incrementato in alcuni casi il divario nei confronti di studenti con disabilità e DSA, che rischiano di essere esclusi dai benefici dell'innovazione. Una delle risorse didattiche digitali maggiormente utilizzata e diffusa è sicuramente il video della registrazione delle lezioni. Questa risorsa didattica digitale però presenta diversi problemi di accessibilità: persone con disabilità uditive non accedono all'audio, persone con disabilità visive non accedono ai contenuti video, persone con disabilità motorie e con DSA hanno difficoltà a prendere appunti. Per garantire l'accessibilità e la completa fruizione di tale risorsa servono trascrizione (Armano et al., 2020) e sottotitoli adeguati. Attualmente ci sono molti sistemi per la trascrizione e la sottotitolazione automatica, con buone performance e in continuo e veloce miglioramento: molti atenei forniscono sistemi e servizi per la trascrizione e la sottotitolazione anche in tempo reale. Tuttavia, nel caso di "lezioni con formule" la trascrizione del parlato in linguaggio naturale non risulta adeguata: risulta, infatti, ambigua, prolissa, non navigabile. È necessario che le formule siano codificate con un linguaggio opportuno, come LaTeX o MathML (Armano et al., 2021).

1. Il progetto VoiceMath

Lo sviluppo di uno strumento informatico per la trascrizione di “lezioni con formule” è stato possibile solo grazie all'evoluzione dei sistemi di speech-to-text e di riconoscimento ottico delle formule. Nel 2021, grazie a un finanziamento della Fondazione CRT e del Dipartimento di Matematica dell'Università di Torino, il Laboratorio Polin ha avviato il progetto Voice-Math con Maize srl (Amore et al, 2021). Il prototipo del software, disponibile a fine 2021, per trascrizioni integrate con LaTeX, è risultato promettente. Quindi la Direzione Sistemi Informativi Portale e E-learning dell'Università di Torino ha deciso di finanziarne l'integrazione con i servizi di ateneo insieme allo sviluppo di altre funzionalità. A ottobre 2022 l'applicativo è stato reso disponibile a un gruppo di docenti per una fase di test e, in seguito alle recenti evoluzioni nell'ambito dei Large Language Model, sono state implementate nuove funzionalità avanzate.

1.1 Caratteristiche tecniche

VoiceMath ha come obiettivo la produzione automatica di un flusso ordinato di contenuti in un documento LaTeX, a partire da (1) videoregistrazione di una lezione di discipline STEM e (2) corrispondenti immagini contenenti il testo scritto a mano dal docente. L'audio del docente, isolato dalla videoregistrazione, viene sottoposto a trascrizione automatica mediante tecnologie di Natural Language Processing (Jurafsky & Manning, 2012) e di Speech-to-Text.

Le immagini, solitamente ricavate dal pdf prodotto dal docente durante la videolezione, vengono sottoposte ad una componente di OCR ed Image-to-LaTeX, ovvero estrazione automatica delle formule contenute all'interno delle immagini, e trasformazione in linguaggio LaTeX.

Le trascrizioni dell'audio e delle immagini sono consultabili tramite un'applicazione web dove l'utente può rivedere/modificare le trascrizioni prodotte dai componenti automatici e combinarle per produrre un solo out-put in formato LaTeX. Il software fornisce alcuni strumenti che affiancano e aiutano l'utente in questa attività: modelli di classificazione e similarità fra testo e formule, player per la navigazione dell'audio estratto dalla videolezione, un ambiente di anteprima e modifica delle formule LaTeX, funzionalità di editing avanzate (Figura 1).

Fig. 1
Screenshot
dell'applicazione
VoiceMath

The screenshot displays the VoiceMath application interface. On the left, a transcript of a lecture is visible, containing text such as: "Proviamo iniziare prima di iniziare la lezione vera e propria. Vorrei solo rievocare una cosa che riguarda quella piccola indagine a cui gentilmente avete risposto e a una persona che però non posso sapere se è in aula o no che ha fatto la laurea triennale in matematica finanziaria, non dalla facoltà che fate e lei, ah scusi niente ha chiesto quel quella mail in cui le ho... Comunque ho risposto, insomma dicendo che non oggi ma dalla settimana prossima entravano in gioco dei discorsi riguardanti un capitolo che si fa nella triennale in matematica che si chiama teoria della misura integrabile di 1900, ah bene, comunque lo pensavo e lei me lo rivoltò a tutti noi ha scritto sulla pagina Moodle del corso una risposta di questi argomenti. Sì." Below this, there are several numbered points (1-10) discussing mathematical topics like "l'ultima cosa che abbiamo fatto", "l'ultima cosa che abbiamo fatto", "l'ultima cosa che abbiamo fatto", "l'ultima cosa che abbiamo fatto", "l'ultima cosa che abbiamo fatto", "l'ultima cosa che abbiamo fatto", "l'ultima cosa che abbiamo fatto", "l'ultima cosa che abbiamo fatto", "l'ultima cosa che abbiamo fatto", "l'ultima cosa che abbiamo fatto".

On the right, there is a section titled "Scegli il tag con cui vuoi copiare le formule:" with a dropdown menu set to "L-1". Below this, there is a list of detected mathematical formulas, each with a circular icon and a "Preview" button. The formulas shown are:

- $f: X \rightarrow y \quad x_0 \in X$
- $f \in C^n(x)$
- $(k2)(df)(x_0)$
- $\exists U(x_0) \quad \exists V(f(x_0))$
- $f: U \rightarrow V$

Principali componenti del software sono i modelli di classificazione e di similarità delle formule, che sono stati aggiornati nel 2024 in relazione all'evoluzione recente delle tecnologie utilizzate (LLM in particolare). La prima componente ha l'obiettivo di identificare le porzioni di testo ad alto contenuto matematico all'interno della trascrizione automatica. L'identificazione si basa su un modello di classificazione binaria, addestrato su dati di dominio, che analizza il testo e classifica gli span come "formula/non formula". La soluzione proposta si basa su un fine-tuning di un modello sentence transformer multilingua, basato su BERT, utilizzando l'approccio SetFit, un framework per il few-shot fine-tuning. Il dataset è composto da 13 testi di trascrizioni di lezioni matematiche, trascritte e annotate manualmente assieme al Dipartimento di Matematica; per l'annotazione si è utilizzato il tag <inizio formula>/<fine formula> per identificare gli span di testo ad alto contenuto matematico. Il dataset così composto è stato preprocessato e diviso in training set, validation set e test set. Sono stati eseguiti due diversi fine-tuning: il primo attuando un bilanciamento del dataset tra le due classi, il secondo senza applicare il bilanciamento. La seconda componente ha l'obiettivo di aiutare l'utente nella revisione delle porzioni ad alto contenuto matematico, suggerendo la possibile formula con cui sostituire il testo evidenziato. L'identificazione si basa su un modello che calcola la distanza tra la porzione di testo e la formula trascritta in LaTeX. Le porzioni di testo classificate da un modello precedente vengono ri-analizzate e preprocessate: nella fase di preprocessing vengono rivisti e parsificati alcuni termini specifici del dominio (ad esempio, "meno" in < - >) per trasformare il testo da linguaggio naturale alla codifica LaTeX della specifica formula. Il modello sviluppato lavora sul testo parsificato e sulle formule trascritte in LaTeX, restituendo una metrica di somiglianza tra i due oggetti in input che permette, quindi, di collegare automaticamente porzioni di testo e formule. Queste due componenti consentono di mettere a disposizione degli utenti la nuova funzionalità di suggerer, che permette all'utente di individuare più facilmente le porzioni di testo da sostituire con codice LaTeX e di collegare la relativa formula convertita da immagine a LaTeX.

1.2 Dal progetto di ricerca al servizio di Ateneo

In seguito all'analisi del prototipo realizzato nel 2021, la Direzione Sistemi Informativi, Portale, E-Learning ha ritenuto opportuno proporre VoiceMath come servizio di ateneo. Si è resa quindi necessaria la definizione di possibili scenari per rendere la soluzione scalabile e sostenibile nel tempo. Fondamentale in questo senso è stato lo sviluppo di funzionalità relative alla gestione multi-utente, il supporto per una coda di elaborazioni e l'integrazione con il sistema single sign-on di Ateneo. L'intera infrastruttura è stata inoltre potenziata per supportare un bacino di utenti più ampio rispetto a quello coinvolto nella fase di prototipo.

In questa fase del progetto è stata predisposta anche una filiera di supporto analoga a quella degli altri servizi online gestiti dalla Direzione SIPE, che prevede la predisposizione di manualistica, un canale di contatto e il censimento degli utenti abilitati all'applicativo, in modo da raggiungerli per comunicazioni di servizio (manutenzioni, disservizi, ecc.).

Il progetto è quindi diventato servizio di Ateneo a supporto dell'obiettivo del Piano Strate-

gico 2021-2026 rispetto all'inclusività (azione 1.2.3 Individuare soluzioni per supportare soggetti in situazioni particolari di disagio e offrire loro le migliori condizioni di studio/lavoro).

Un riferimento al servizio è stato anche incluso negli Obiettivi di Accessibilità (rif. art. 9, comma 7, del D.L. 18 ottobre 2012, n. 179) per l'Ateneo per l'anno 2022.

Riferimenti bibliografici

Amore M., Armano T., Bernareggi C., Capietto A., Coriasco S., Crespan R., Duc-ci M., Gabrielli M. L., Guadalupi M. F., Mazzei A., Mazzei A., Sofia A. Tarasconi F. (2021), Riconoscimento vocale di formule, Conferenza Garr 2021.

Armano T., Baratto G., Bernareggi C., Bosco M., Capietto A., Caramagna M., Ducci M., Mazzei A., Mazzei A., Tarasconi F. (2021), Matematica a voce parte 1: trascrizione di lezioni di matematica, Atti del MoodleMoot Italia 2020, (2021), pp 23-27.

Armano T., Capietto A., Ahmetovic D., Bernareggi C., Coriasco S., Ducci M., Magosso C., Mazzei A., Murru N., Sofia A. (2020), Accessibilità di contenuti digitali per le STEM: un problema aperto. Alcune soluzioni inclusive per l'accessibilità di formule e grafici, Mondo Digitale, (vol. 89), pp 1-12.

Jurafsky, D., & Manning, C. (2012), Natural language processing. Instructor, 212(998), 3482.

Autori

Tiziana Armano tiziana.armano@unito.it

Tecnico della ricerca presso il Dipartimento di Matematica "G. Peano" dell'Università degli Studi di Torino. Si occupa di e-learning e tecnologie per la didattica. Ha fatto parte del gruppo di ricerca del progetto "Per una matematica accessibile e inclusiva" dal 2012 ed è membro del Laboratorio Polin. In questo ambito segue in particolare i progetti VoiceMath, SpeechMatE e Audiofunctions.web.

Gabriele Baratto gabriele.baratto@unito.it

Attualmente membro dello staff E-learning, Collaboration online e Architetture dell'Università di Torino. Sempre presso l'Ateneo torinese ha conseguito la laurea specialistica in Scienze Linguistiche, con una tesi in sociolinguistica, e il dottorato di ricerca in Scienze del Linguaggio e della Comunicazione, con una tesi in linguistica italiana. È stato collaboratore dell'Atlante Linguistico ed Etnografico del Piemonte Occidentale (ALEPO), docente a contratto presso l'Università di Torino, docente di scuola secondaria e coordinatore di diversi progetti per contrastare l'insuccesso scolastico.

Cristian Bernareggi cristian.bernareggi@unimi.it

Dottorato di ricerca in informatica nel 2006. Tra il 2004 e il 2009 ha condotto ricerca in ambito HCI nei progetti europei LAMBDA ed @Science. Dal 2023 è tecnico della ricerca presso il Dipartimento di Matematica "G. Peano" dell'Università degli Studi di Torino. Tra il 2011 e il 2018 ha partecipato alla fondazione e alla crescita dello spin-off EveryWare Technologies per lo sviluppo di tecnologie assistive su dispositivi mobili. Dal 2018 collabora con il Laboratorio Polin sull'accessibilità alle STEM per persone con disabilità.

Sara Bertoldo sara.bertoldo@maize.io

Lavora come linguista computazionale e sviluppatrice in Maize. Laureata in Lingue, civiltà e scienze del linguaggio presso l'università Ca' Foscari di Venezia e successivamente in Informatica Umanistica presso l'università di Pisa. Coniuga queste due anime in progetti di analisi di testo, assistenti conversazionali, generazione e processamento del linguaggio naturale dal forte carattere consulenziale.

Erika Brunetto erika.brunetto@unito.it

Laureata magistrale in Matematica presso l'Università degli Studi di Genova. Dopo la laurea ha avuto diverse esperienze sia in ambito didattico che aziendale. Dopo un assegno di ricerca di due anni presso CNR-IMATI (Istituto di Matematica Applicata e Tecnologie Informatiche) della sede di Genova, da più di un anno collabora con il Laboratorio Polin di Unito come borsista di ricerca.

Anna Capietto anna.capietto@unito.it

Professore Ordinario di Analisi Matematica. Da 15 anni è Referente per le tematiche degli studenti con disabilità presso il Dipartimento di Matematica. Dal 2013 è responsabile del progetto "Per una matematica accessibile e inclusiva", le cui attività hanno condotto nel novembre 2018 all'istituzione del Laboratorio per la ricerca e la sperimentazione di nuove tecnologie assistive per le STEM "S.Polin".

Manuela Caramagna manuela.caramagna@unito.it

Responsabile dello Staff E-Learning, Collaboration on line, Architetture presso la Direzione Sistemi Informativi, Portale, E-learning dell'Università di Torino. Laureata in Comunicazione Multimediale e di Massa e con un Master in IT Governance and Compliance, si occupa dal 2006 di gestione di servizi web e di tecnologie per la didattica.

Sandro Coriasco sandro.coriasco@unito.it

Professore Associato di Analisi Matematica presso il Dipartimento di Matematica dell'Università di Torino. Referente per l'inclusione di studenti e studentesse con DSA. Si occupa di analisi di Fourier e di analisi globale su varietà non compatte e su varietà con singolarità. È un componente del gruppo di ricerca sulle tecnologie assistive presso il Laboratorio "S. Polin". In tale ambito si occupa, in particolare, dello sviluppo del pacchetto LaTeX accessibility e, più in generale, dei temi collegati all'accessibilità dei contenuti matematici.

Nazareno Defrancesco nazareno.defrancesco@maize.io

Deep Learning e Software Analyst presso Maize srl.

Mattia Ducci mattia.ducci@unito.it

Laurea magistrale in informatica all'Università degli Studi di Milano nel 2019. Durante e dopo l'università ha collaborato a vari progetti incentrati sul tema dell'accessibilità: da soluzioni su dispositivi mobili per persone con disabilità visive (Università degli Studi di Milano, Università degli Studi di Pavia) a soluzioni per rendere più accessibili le materie scientifiche (Università degli Studi di Torino). Ha lavorato dal 2018 al 2023 come consulente informatico in un'azienda per sviluppo mobile iOS, cross-platform e full-stack e ora lavora come libero professionista su temi legati

all'accessibilità.

Maria Luisa Gabrielli marialuisa.gabrielli@maize.io

Lavora come Data Scientist presso Maize. Laureata in Matematica presso l'Università di Torino, si è specializzata in Analisi numerica, in particolare avvicinandosi al mondo dell'elaborazione e del riconoscimento di immagini. Ha esperienza in Data Analysis, nell'applicazione di algoritmi di Machine Learning e Deep Learning e in tecniche di Computer Vision.

Mariafrancesca Guadalupi mariafrancesca.guadalupi@maize.io

Lavora come linguista computazionale dopo aver conseguito una laurea magistrale in Linguistica e Comunicazione Persuasiva, Tecnologia e Studi Cognitivi ed un Master in Disability Management. Si occupa di progetti riguardanti: text classification, assistenti conversazionali, sentiment analysis ed opinion mining. Sta studiando per approfondire il mondo delle tecnologie assistive.

Alessandro Mazzei alessandro.mazzei@unito.it

Laurea in Fisica (indirizzo Cibernetico) presso l'Università Federico II di Napoli nel 2000. Dottore di Ricerca in Informatica presso l'Università degli Studi di Torino nel 2006, dove è Professore Associato presso il Dipartimento di Informatica. I suoi interessi di ricerca includono le ontologie giuridiche, la sintassi e la semantica computazionale, la generazione automatica del linguaggio, i sistemi di dialogo, il ragionamento automatico, le tecnologie assistive e il food computing. È autore di oltre 100 lavori scientifici peer reviewed.

Angelo Sacca angelo.sacca@unito.it

Direttore della Direzione Sistemi Informativi, Portale, E-learning dell'Università di Torino. Presidente dell'Associazione Italiana Comunicatori d'Università (AICUN).

Francesco Tarasconi francesco.tarasconi@maize.io

Data Scientist and Artificial Intelligence Manager in MAIZE. In MAIZE si occupa di Natural Language Processing, AI generativa e sistemi di supporto alle decisioni. Ha supportato ENI nella realizzazione del Progetto di AI Safety Pre-Sense per la tutela dell'integrità fisica dei dipendenti. Collabora con l'Università di Torino nello sviluppo di software con l'obiettivo di migliorare l'accessibilità alle lezioni di materie STEM. Conduce attività di ricerca industriale in collaborazione con il Politecnico di Torino su temi di AI applicata all'ambito giuridico.