

DICHIARAZIONI SOSTITUTIVE DI CERTIFICAZIONI

(art. 46 D.P.R. n. 445/2000)

DICHIARAZIONI SOSTITUTIVE DELL'ATTO DI NOTORIETÀ

(art. 47 D.P.R. n. 445/2000)

Il sottoscritto

COGNOME ALESSIO

(per le donne indicare il cognome da nubile)

NOME MARCO**NATO A****PROV.****IL****ATTUALMENTE RESIDENTE A****PROV.****INDIRIZZO****TELEFONO**

Visto il D.P.R. 28 dicembre 2000, n. 445 concernente "T.U. delle disposizioni legislative e regolamentari in materia di documentazione amministrativa" e successive modifiche ed integrazioni;

Vista la Legge 12 novembre 2011, n. 183 ed in particolare l'art. 15 concernente le nuove disposizioni in materia di certificati e dichiarazioni sostitutive (*);

Consapevole che, ai sensi dell'art.76 del DPR 445/2000, le dichiarazioni mendaci, la falsità negli atti e l'uso di atti falsi sono punite ai sensi del Codice penale e delle leggi speciali vigenti in materia, dichiara sotto la propria responsabilità:

che quanto dichiarato nel seguente curriculum vitae et studiorum
comprensivo delle informazioni sulla produzione scientifica
corrisponde a verità

Curriculum vitae et studiorum**MAG. 2023 – SET. 2023: COLLABORAZIONE DI RICERCA POST-LAUREA MAGISTRALE****[1] – PUBBLICAZIONE SU CONFERENZA INTERNAZIONALE****AUTORE**

Marco Alessio, Guglielmo Faggioli, Nicola Ferro

TITOLO

DECAF: a Modular and Extensible Conversational Search Framework

CONFERENZASIGIR 2023

28 / 09 / 2023

MARCO ALESSIO

LINK <https://doi.org/10.1145/3539618.3591913>
DATA Luglio 2023



[2] – SOFTWARE PUBBLICATO SU REPOSITORY OPEN ACCESS

AUTORE Marco Alessio, Guglielmo Faggioli, Nicola Ferro
REPOSITORY <https://github.com/alemarco96/DECAF>
LICENZA [Creative Commons Attribution-ShareAlike 4.0 International License](#)
DATA Luglio 2023
NOTE Risorsa Software Open-Source, relativa alla pubblicazione [1]



[3] – PUBBLICAZIONE SU CONFERENZA NAZIONALE

AUTORE Marco Alessio, Guglielmo Faggioli, Nicola Ferro
TITOLO A Modular Framework for Conversational Search Reproducible Experimentation
CONFERENZA [SEBD 2023](#)
LINK <https://ceur-ws.org/Vol-3478/paper16.pdf>
DATA Luglio 2023
NOTE Extended Abstract della pubblicazione [1]

La collaborazione con il gruppo di ricerca del Prof. Nicola Ferro (vedi [5] per maggiori dettagli) prosegue anche nel periodo post-laurea magistrale. Il progetto di tesi magistrale relativo al *framework* DECAF è stato rielaborato per produrre un Resource Paper, che è stato sottomesso alla conferenza internazionale SIGIR edizione 2023. Contestualmente alla pubblicazione del paper, la risorsa DECAF è stata resa disponibile su un repository open access: sono liberamente consultabili il codice sorgente, la documentazione, le *run* generate da diverse *pipeline* oggetto di valutazione nel paper, oltre a guide e tutorial necessarie per la corretta installazione e configurazione del *framework*, per replicare tutti i risultati mostrati nel paper, e per lo svolgimento di nuovi esperimenti. Inoltre, un *extended abstract* del Resource Paper è stato pubblicato alla conferenza italiana SEBD edizione 2023.

DIC. 2019 – APR. 2023: LAUREA MAGISTRALE



[4] – LAUREA MAGISTRALE

TITOLO [LM-32 – Laurea Magistrale in Ingegneria Informatica](#)
RILASCIATO DA Università degli Studi di Padova
VALUTAZIONE 110 / 110
DATA 01 Dicembre 2019 – 13 Aprile 2023

Nel periodo compreso tra Dicembre 2019 ed Aprile 2023 ho studiato presso l'Università degli Studi di Padova per conseguire la laurea magistrale in Ingegneria Informatica.

SET. 2022 – APR. 2023: TESI DI LAUREA MAGISTRALE



[5] – TESI DI LAUREA MAGISTRALE

AUTORE	Marco Alessio
TITOLO	Design and Development of an Extensible and Configurable Framework for Conversational Search Experiments
LINK	https://thesis.unipd.it/handle/20.500.12608/45641
DATA	Settembre 2022 – 13 Aprile 2023

A partire da Settembre 2022, ho svolto la tesi magistrale sotto la supervisione del Prof. Nicola Ferro nel ruolo di relatore e del Dott. Guglielmo Faggioli come co-relatore, entrambi afferenti all'Intelligent Interactive Information Access Hub del Dipartimento di Ingegneria dell'Informazione (DEI) dell'Università di Padova. L'obiettivo della tesi consiste nel progettare e sviluppare DECAF, un *framework* modulare ed estensibile che permetta una rapida prototipazione e sviluppo di agenti di ricerca conversazionale. Il paradigma di ricerca conversazionale prevede un'intuitiva ed efficace interazione tra l'utente ed il sistema attraverso lo scambio di frasi in linguaggio naturale, e risulta essere sempre più adottato in svariati scenari applicativi. La recente diffusione capillare dei sistemi conversazionali ha portato alla nascita di una moltitudine di sistemi con implementazioni *custom* e varianti dei medesimi modelli di *Information Retrieval*, esacerbando la crisi di riproducibilità già osservata in quest'area di ricerca. Il contributo portato da DECAF consiste nell'identificazione dei principali moduli costituenti un sistema conversazionale, nella definizione ed implementazione di un *framework* modulare, e nella possibilità di svolgere esperimenti caratterizzati da un alto grado di riproducibilità. DECAF integra tutti i principali moduli che caratterizzano i moderni sistemi conversazionali stato dell'arte, come illustrato dalla Figura 1, e consente l'integrazione di tecniche basate su *Machine Learning* e *Large Language Models*. Il *framework* include numerosi componenti stato dell'arte pronti all'uso per l'utilizzatore, quali diverse tecniche di riscrittura delle *query*, *scoring functions* basate su rappresentazioni sia sparse (*bag of words*) che dense dei documenti, e funzioni di *re-ranking*. Si noti che DECAF permette inoltre di eseguire la fase di indicizzazione del *corpus* di documenti.

Per svolgere l'attività di tesi magistrale, ho collaborato in presenza con il gruppo di ricerca di Nicola Ferro e Guglielmo Faggioli recandomi ogni giorno presso il loro laboratorio all'Università di Padova. A livello metodologico, la tesi magistrale è stata realizzata secondo quattro *step* differenti. La prima attività ha consistito nello studio della letteratura scientifica relativa ai sistemi conversazionali, con lo scopo di acquisire conoscenze nell'ambito e di determinare quali siano i principali moduli che costituiscono tutti i moderni sistemi conversazionali allo stato dell'arte. Nella seconda fase ho proceduto ad una prima implementazione del *core* di DECAF, poi seguito dall'integrazione di nuovi componenti a mano a mano che sperimentavo con nuove tecniche, algoritmi, e modelli. Il *core* del *framework* è sviluppato in Java, garantendo la compatibilità con le principali librerie di *Information Retrieval* quali Apache Lucene, Terrier e Anserini, e permette l'integrazione di script Python, supportando così anche moderne tecniche basate su *Machine Learning* e *Large Language*

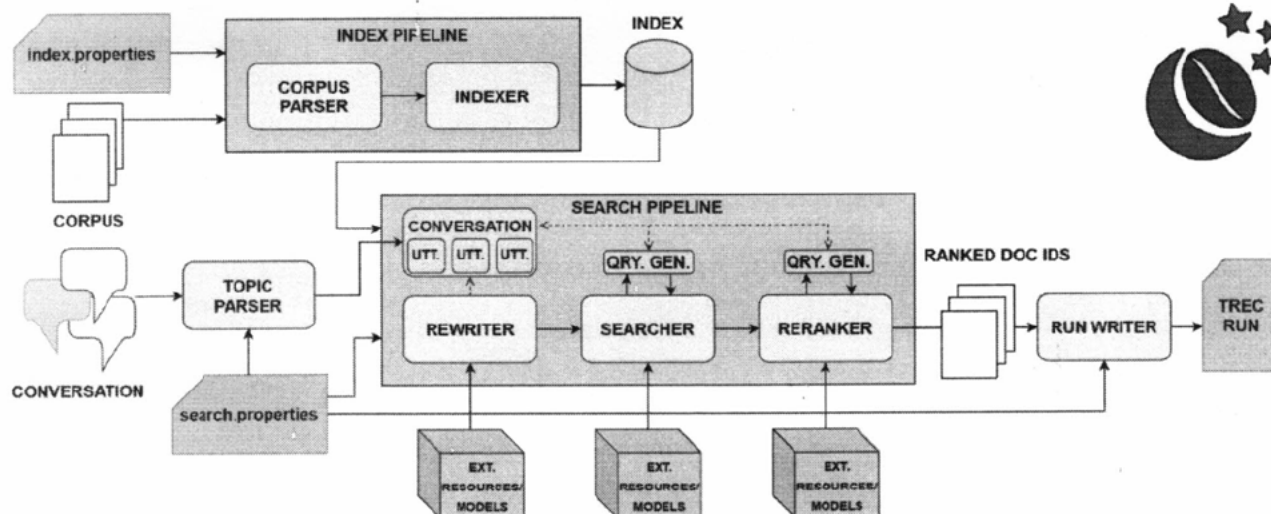


Figura 1: Architettura di DECAF

Models. Attraverso la combinazione di diversi componenti scelti per ciascun modulo, DECAF permette di istanziare molteplici *pipeline* diverse. Per valutarne la diversa efficacia, vengono impiegate due ben note collezioni di test conversazionali come TREC CAsT 2019 e TREC CAsT 2020. Nella terza fase, DECAF è stato dotato di un sistema per poter istanziare a piacimento i diversi componenti esclusivamente attraverso l'utilizzo di file di configurazione testuali, senza richiedere modifiche al codice e di ricompilare l'intero *framework*. Inoltre, il codice di DECAF è stato consolidato ed opportunamente documentato. Infine, ho provveduto alla scrittura vera e propria della tesi.

MAR. 2021 – SET. 2021: PROGETTO DI GRUPPO E PUBBLICAZIONE PAPER SU CLEF 2021



[6] - PUBBLICAZIONE SU CONFERENZA INTERNAZIONALE

AUTORE	Marco Alessio, Nicola Levorato, Edoardo Raimondi
TITOLO	A Search Engine System for Touché Argument Retrieval task to answer Controversial Questions
CONFERENZA	CLEF 2021
LINK	https://ir.webis.de/anthology/2021.clef_conference-2021w.217/
DATA	Settembre 2021

Durante il ciclo di studi per il conseguimento della laurea magistrale, nel periodo compreso tra Marzo e Settembre 2021, ho svolto un progetto di gruppo per il corso di “Reperimento dell’Informazione” che ha portato alla pubblicazione di un paper sulla conferenza internazionale CLEF edizione 2021. Il mio gruppo ha partecipato al Task 1 dello *shared task* “Touché: Argument Retrieval”. Lo scopo consiste nel supportare l’utente del sistema che ricerca informazioni e ragionamenti attinenti da usare per rispondere a domande relative ad argomenti controversi. Per questo *task* è importante recuperare documenti che, oltre ad essere rilevanti per la *query*, siano anche “ben scritti”. Infatti, vengono rilasciati due distinti *relevance judgement* in quanto gli

annotatori umani giudicano ciascun documento due volte: la prima prende in esame esclusivamente la sua rilevanza, mentre la seconda considera anche la sua qualità retorica. Il *corpus* di documenti utilizzato è “args.me”, composto da oltre 380 mila dibattiti estrapolati da 5 siti web di dibattito pubblico in lingua inglese. La valutazione finale consiste nel calcolo del valore medio della metrica *Normalized Discounted Cumulative Gain* a *cutoff* 5 (nDCG@5) per tutti i 50 *topics* ottenuto dalla migliore *run* di ciascun partecipante.

A livello metodologico, il sistema del mio gruppo è stato sviluppato in Java utilizzando la libreria Apache Lucene versione 8.8.1 come fondamento per l'intero progetto. La fase di analisi testuale prevede inizialmente di eliminare dati superflui, come i *link*, attraverso l'uso di un'espressione regolare, poi il testo viene suddiviso nelle singole parole che lo compongono (*tokenization*), infine vengono scartate molte parole di uso comune in lingua inglese attraverso l'utilizzo di una *stop list* creata ad-hoc per questo progetto. Si noti che non viene impiegato nessun algoritmo di *stemming*. Quindi viene eseguita una fase di *query expansion*, attraverso l'ausilio di una versione *custom* del dizionario Word-Net 3.0. Per ogni parola presente nel *topic*, vengono aggiunte forme simili (come, ad esempio, i plurali dei nomi o diverse coniugazioni dei verbi), sinonimi e contrari. A questo punto è possibile costruire la *query* che Lucene utilizza per recuperare la lista dei documenti più rilevanti, usando BM25 come funzione di *scoring*. Infine, avviene una fase di *re-ranking*: un modello di regressione lineare valuta la qualità generale della scrittura del testo di ciascun documento. Gli *score* di rilevanza e qualità sono min-max normalizzati, quindi combinati linearmente per ottenere gli *score* finali.

AGO. 2015 – NOV. 2019: LAUREA



[7] – LAUREA

TITOLO L-8 – Laurea in Ingegneria Informatica
RILASCIATO DA Università degli Studi di Padova
VALUTAZIONE 98 / 110
DATA 26 Agosto 2015 – 18 Novembre 2019



[8] – CERTIFICAZIONE INGLESE B2 – READING AND LISTENING

TITOLO English B2 - Reading and Listening
RILASCIATO DA Università degli Studi di Padova
DATA 05 Febbraio 2019

Durante il periodo compreso tra Agosto 2015 e Novembre 2019 ho studiato presso l'Università degli Studi di Padova per conseguire la laurea in Ingegneria Informatica. Nel corso di questo ciclo di studi, ho ottenuto la certificazione linguistica “B2 Reading and Listening” in lingua inglese, rilasciata in seguito al superamento della prova di accertamento dei requisiti sostenuta presso il Centro Linguistico di Ateneo.

COMPETENZE

**[9] – LINGUAGGI DI PROGRAMMAZIONE**

AVANZATO	Java
INTERMEDIO	Python
INTERMEDIO	C / C++
BASE	R

**[10] – LIBRERIE DI INFORMATION RETRIEVAL**

INTERMEDIO	Apache Lucene
BASE	PyTerrier

(*) ai sensi dell'art. 15, comma 1 della Legge 12/11/2011, n. 183 le certificazioni rilasciate dalla P.A. in ordine a stati, qualità personali e fatti sono valide e utilizzabili solo nei rapporti tra privati; nei rapporti con gli Organi della Pubblica Amministrazione e i gestori di pubblici servizi, i certificati sono sempre sostituiti dalle dichiarazioni sostitutive di certificazione o dall'atto di notorietà di cui agli artt. 46 e 47 del DPR 445/2000