UNIVERSITÀ DEGLI STUDI DI SALERNO

Dipartimento di Informatica

Corso di Laurea Triennale in Informatica

TESI DI LAUREA

ShallWeGo: AI-assisted mobility crowdsourcing platform

RELATORE

Prof. **Fabio Palomba**

Università degli studi di Salerno

CANDIDATO

Hermann Senatore

Matricola: 0512105743

Anno Accademico 2020-2021

Il mondo è come un libro e chi non viaggia ne conosce una pagina soltanto.

(Sant'Agostino)

Sommario

La Tesi sviluppata si posiziona nell'ambito della creazione di strumenti informatici per assistere gli utenti del Trasporto Pubblico Locale nella loro esperienza quotidiana. L'obiettivo principale di questa Tesi consiste nello sviluppo di un'applicativo mobile che permetta l'accesso ad una piattaforma online chiamata ShallWeGo basata sulla collaborazione tra utenti che ha come scopo quello di consentire lo scambio di informazioni sull'organizzazione del Trasporto Pubblico Locale in una determinata zona. Il vero motore di ShallWeGo è quindi il singolo utente, che può mettere a disposizione la propria conoscenza sul topic in questione (organizzazione delle fermate sul territorio, aziende di trasporto e linee espletate da queste ultime) agli altri utenti, che quindi possono sopperire alle potenziali difficoltà di comunicazione da parte delle aziende di trasporto. Particolare attenzione deve essere dedicata tuttavia all'affidabilità di queste segnalazioni che devono essere controllate in qualche modo. È stato quindi previsto, tramite la definizione di un Agente Intelligente, un sistema di verifica delle segnalazioni anch'esso basato sulla partecipazione attiva degli utenti della piattaforma. Data una segnalazione sarà quindi possibile andare a determinare il gruppo di utenti che secondo determinate "metriche" risultino più adatti a verificarla e stabilire se quest'ultima possa essere integrata o meno sulla piattaforma.

Indice

Elenco delle figure

Elenco delle tabelle

CAPITOLO 1

Introduzione

1.1 Contesto applicativo e Motivazioni

1.1.1 Lo stato del Trasporto Pubblico Locale

Negli ultimi due anni, complici anche gli avvenimenti che stanno interessando il mondo ed i successivi provvedimenti, le abitudini dei cittadini in tema di mobilità sono drasticamente cambiate. Tra i numerosissimi settori che sono interessati da questa ondata di cambiamenti ce n'è uno in particolare che ne ha risentito particolarmente: quello del Trasporto Pubblico Locale (**TPL**) che, secondo un rapporto stilato da varie associazioni di rappresentanza di aziende di trasporto presentato al Parlamento Italiano all'inizio del 2021 (Asstra [2021]) e ripreso dal quotidiano "Il Sole 24 Ore" in data 25 Gennaio dello stesso anno De Girolamo [2021] ha visto un crollo dei ricavi dell'entità dei Miliardi di Euro. Nello specifico, dal crollo della fruizione dei mezzi pubblici si stima che nelle casse delle aziende siano entrati complessivamente 250 milioni di Euro in meno al mese, portando le perdite complessive a circa 2 Miliardi. Tutto ciò, come anticipato, è causato sostanzialmente da un crollo delle presenza da parte dei cittadini su autobus, treni e metropolitane in un primo momento dovuto alle restrizioni imposte, almeno nel caso dell'Italia, nella prima metà del 2020 e complice la percezione da parte del cittadino dei mezzi pubblici come "poco sicuri" dal punto di vista sanitario risultando, continua il rapporto, in un calo del 90% della domanda. Questi due fattori (crollo della domanda e conseguente diminuzione delle entrate) hanno generato in molti casi conseguenze pesanti sullo "stato di salute" delle aziende del settore, con successive

rimodulazioni orarie e una sostanziale diminuzione delle corse, specie nelle prime e nelle ultime ore della giornata. La diminuzione delle corse inevitabilmente apre la porta ad uno dei problemi principali che sta affliggendo il settore nell'ultimo periodo: l'affollamento delle corse stesse.

1.1.2 Tecnologia e TPL

Con la sempre più capillare diffusione della tecnologia nelle mani del cittadino da un decennio a questa parte, i principali attori che operano in questo campo hanno riconosciuto l'importanza di andare a creare un sistema di informazione puntuale sul funzionamento del TPL in una determinata zona. Piattaforme come Google Maps, ad esempio, già da tempo forniscono dettagli sulla presenza di fermate e dettagli delle corse delle aziende di trasporto che operano in quella zona. Nel caso particolare di Google Maps, questi dati, secondo la documentazione ufficiale del servizio "Google Transit" disponibile all'indirizzo `https://support.google.com/transitpartners/answer/1111471` sono ricavati principalmente dalle comunicazioni che giungono alla piattaforma da parte delle singole aziende che decidono di partecipare al programma, avendo poi l'opzione di condividere anche dati in tempo reale che permettano agli utenti di seguire in diretta l'andamento delle corse delle linee che gli interessano, tramite la piattaforma "Realtime Transit", che comunque è riservato alle aziende di trasporto.

1.1.3 Problemi delle attuali soluzioni

Come accennato in precedenza, i dati che sono disponibili sulle piattaforme più diffuse sono comunicati periodicamente dalle aziende che aderiscono a programmi come il sopracitato Google Transit su Maps. Tuttavia, questo approccio non è esente da problemi di natura pratica: la presenza dei dati disponibili sulle varie piattaforme risulta quindi, nella maggior parte dei casi, essere subordinata all'adesione delle varie aziende alle stesse. Se questo problema non si pone per le aziende di medie/grandi dimensioni che hanno a disposizione risorse economiche e tecnologiche, lo stesso non si può dire per le aziende più piccole a carattere locale, che potrebbero avere difficoltà di carattere logistico anche solo nel posizionare gli adeguati segnali di riconoscimento delle fermate che utilizzano. Ne consegue quindi che gli utenti occasionali di queste aziende di trasporto (si pensi ad esempio ad una persona che si trova per la prima volta in una determinata zona) siano costretti a cercare informazioni presso altri cittadini che "ne sanno più di loro". Una piattaforma che ha provato a porre rime-

dio a questo problema è Moovit che ha lanciato nel 2015 un servizio chiamato "Mooviters' Community" che permette agli utenti di andare ad aggiungere dettagli alla piattaforma stessa che non siano stati già comunicati alle aziende. Questo approccio verrà trattato nel capitolo successivo dedicato proprio alle piattaforme presenti ad oggi sul mercato.

1.2 Obiettivi della tesi

L'obiettivo di questo lavoro è quindi proporre un primo esempio di soluzione ai due problemi di cui si è trattato in precedenza: creare una piattaforma, chiamata **ShallWeGo**, che metta a disposizione dei suoi membri gli strumenti adatti a diffondere la propria conoscenza in merito di organizzazione di linee e fermate nella loro zona con gli altri utenti, formando quindi una community che vede nella reciproca collaborazione il principale mezzo per raggiungere una vera e propria utilità per il singolo cittadino.

1.3 Metodologie e risultati

Il risultato di questa esperienza di Tesi è stato quindi quello di aver sviluppato una prima versione di un applicativo mobile e della relativa componente server che implementasse l'idea di cui si è discusso. L'applicativo consiste quindi in un semplice file .apk per Android installabile e che, attualmente, previa registrazione ed ottenimento di uno username, permetta a chi vuole di cominciare ad effettuare segnalazioni riguardanti fermate, linee ed aziende di trasporti che conosce e che sa operare in una determinata zona. È, inoltre, in sviluppo una funzionalità che preveda la possibilità da parte di un utente di comunicare in tempo reale la propria posizione e, contemporaneamente, specificare su che linea si trova, così da permettere ad altri utenti interessati di farsi un'idea di quanto debbano aspettare per salire a bordo.

1.3.1 Crowdsourcing

La piattaforma in sviluppo basa il suo funzionamento sulla tecnica del crowdsourcing, ovvero sul recuperare i dati di interesse direttamente da persone sul campo e che ritengano di possedere un'informazione che potrebbe tornare utile alla comunità. Tra queste:

- Posizione di una fermata in un determinato punto
- "Equipaggiamento" di una fermata (in termini di pensilina, segni identificativi e quadri orari disponibili al pubblico)

- Utilizzo di una fermata da parte di una determinata linea
- Destinazioni di una linea
- Eventi transitori (come presenza di traffico, avvisi su deviazioni o strade chiuse)

Questo approccio permette quindi al sistema di essere indipendente da qualsivoglia fonte di informazione o API messa a disposizione pubblicamente da attori terzi (come ad esempio Google stessa o anche Moovit), i cui costi non sempre sono sostenibili sul lungo periodo.

1.3.2 Il problema delle API

Al di là dei dati sul trasporto pubblico locale, l'applicazione oggetto di questo lavoro di Tesi fa largo uso di mappe e dati geografici in generale. Google Maps mette a disposizione le sue API in maniera gratuita fino ad esaurimento di un credito (che viene scalato man mano che si utilizzano i propri servizi) che si rinnova ogni mese. Superata questa soglia, è necessario sostenere un certo prezzo. Le tariffe, consultabili all'indirizzo `https://cloud.google.com/maps-platform/pricing`, prevedono infatti la possibilità di usufruire un credito gratuito dell'entità di 200$ mensili. In particolare, il servizio di Geocoding [1] costa 5$ ogni 1000 richieste. Il prezzo, quindi, aumenta all'aumentare delle richieste che vengono effettuate da chi utilizza l'applicazione. Sono tuttavia disponibili delle alternative di terze parti che permettono l'accesso gratuito a degli endpoint per effettuare le operazioni di Geocoding. Uno dei migliori servizi di questo genere è rappresentato da **Nominatim** che è utilizzato, tra l'altro, anche dal progetto **OpenStreetMap** per implementare la feature di ricerca nella loro piattaforma. Uno dei vantaggi principali di Nominatim consiste nella possibilità da parte di un utente di installare la propria istanza del server, importando all'interno di un database una serie di dati provenienti da diverse fonti. La strada del self-hosting è quella che preferita nel caso di ShallWeGo, poiché mentre disponibile un server pubblico a cui possono essere indirizzate delle richieste, esso impone una restrizione di esattamente una interrogazione al servizio di Geocoding al secondo, pena il *ban* dell'indirizzo IP sorgente dall'utilizzo del servizio. I dati utilizzati da ShallWeGo sono quelli messi a disposizione da Geofabrik [2]. I dati sono disponibili sia per singoli paesi sia per intere regioni o anche addirittura per il mondo intero (che però ha come conseguenza, chiaramente, l'aumento esponenziale della

[1] Ovvero il processo che consente di ottenere il nome logico di un luogo associato ad una coppia di coordinate (Latitudine, Longitudine) e viceversa

[2] `https://download.geofabrik.de/`

dimensione del database ed un allungamento considerevole dei tempi di importazione dei dati).

1.4 Struttura della tesi

La tesi è strutturata principalmente in cinque parti, che coprono gli aspetti principali del lavoro svolto, con un'attenzione particolare alla metodologia utilizzata per implementare in modo efficiente la valutazione delle segnalazioni che pervengono alla piattaforma da parte degli utenti.
In particolare:

- Il Primo Capitolo funge da introduzione al lavoro svolto per realizzare la piattaforma
- Il Secondo Capitolo descrive le principali soluzioni che sono attualmente utilizzabili dagli utenti per quanto riguarda il dominio del problema trattato dal lavoro di Tesi.
- Il Terzo Capitolo descrive accuratamente l'approccio usato per la valutazione delle segnalazioni degli utenti, che fa uso di una tecnica di **Intelligenza Artificiale** che permette di selezionare gli utenti più adatti a valutare una determinata segnalazione
- Il Quarto Capitolo, invece, illustra l'architettura su cui si basa l'applicazione e i dettagli implementatiti in termini di Framework, Linguaggi di Programmazione e Librerie utilizzate per lo sviluppo dell'applicativo Mobile.
- Infine, il Quinto Capitolo riassume tutti gli sviluppi futuri che permetterebbero all'applicazione di uscire dallo stato di "demo" ed essere quindi messa a disposizione del pubblico.

CAPITOLO 2

Stato dell'arte

Lo scopo di questo capitolo è di illustrare la situazione attuale dei vari applicativi che sono a disposizione degli utenti del trasporto pubblico locale (TPL) per permettergli di organizzare i loro spostamenti.

I principali attori sul panorama della diffusione di informazioni sul TPL sono senz'altro la piattaforma Maps di Google ed in particolare Moovit, che ha acquisito una maggiore popolarità nell'ultimo anno.

2.1 Google Maps

Transit è un servizio messo a disposizione da Google ed integrato nella piattaforma Maps che consente all'utente di visualizzare le informazioni *statiche* messe a disposizione dalle varie aziende di trasporto tramite il programma "Google Transit Partners", che quindi consistono nella programmazione giornaliera delle corse e dell'organizzazione delle fermate sul territorio in questione.
Oltre ai dati statici, le aziende di trasporto possono condividere con Google mediante la piattaforma **Realtime Transit** che quindi permette di visualizzare all'interno di Maps i dati in tempo reale sulla posizione dei mezzi pubblici in un determinato istante. Questa funzionalità, così come tutti i dati relativi al trasporto pubblico su Maps sono disponibili solo su iniziativa delle aziende che decidono di partecipare al programma. Un esempio di azienda che mette a disposizione i dati realtime è per esempio l'ATAC di Roma, che li offre a partire da Settembre 2019 (Adnkronos [2019])
I dati (siano essi statici o dinamici) sono comunicati a Google su base regolare dalle aziende (tipicamente settimanale, secondo quanto riportato nella sezione FAQ della pagina dedicata [3]) tramite il formato GTFS *(General Transit Feed Specification)* che permette di strutturare in maniera efficiente i dati riguardanti il trasporto pubblico.

Esistono due "varianti" del formato GTFS:

- **GTFS Statico**, che raccoglie i dati "statici" (quindi gli orari previsti delle corse e l'organizzazione delle fermate sul territorio)
- **GTFS Realtime**, che raccoglie i dati "dinamici" (come l'andamento delle corse in tempo reale)

Tuttavia, come accennato in precedenza, questi dati sono comunicati su richiesta delle aziende e solo un numero limitato di queste ultime (almeno in Italia) aderisce al programma, limitando così i dati a disposizione dell'utente.

[3] `https://developers.google.com/transit/gtfs/guides/faq`

2.2 Moovit ed il crowdsourcing.

Moovit è una piattaforma che al contrario di Google Maps si occupa esclusivamente di mobilità. Nasce nel 2012 e nel 2015 viene acquistata da Intel. Attualmente Moovit si sta impegnando per sviluppare soluzioni di Mobility as a Service. Il Mobility as a Service (MaaS) è un concetto relativamente nuovo, che permette all'utente di scegliere il modo che ritiene più adatto di spostarsi (tramite mezzi pubblici, car sharing e simili) usando una sola piattaforma che mette a disposizione tutti questi servizi, favorendo quindi l'interoperabilità tra i vari mezzi di trasporto.

Dal 2015, inoltre, Moovit mette a disposizione un servizio chiamato **Mooviter Community**, che permette al singolo utente della piattaforma di partecipare alla mappatura delle fermate e delle linee della sua zona tramite un Editor accessibile via web. I cambiamenti dovranno quindi essere approvati dagli amministratori della piattaforma o da utenti esperti.

In particolare, la piattaforma di Moovit organizza gli utenti in "livelli", che potranno essere "scalati" man mano che si matura esperienza in termini di segnalazioni. Nello specifico, se un utente x di livello n effettua una segnalazione, chi potrà verificare questa segnalazione sarà solamente un certo utente z il cui livello risulta almeno $n + 1$ o che sia membro del Team. Il problema di questo approccio è che se un utente con molta esperienza si unisce alla piattaforma in un certo momento, egli non potrà immediatamente cominciare a verificare segnalazioni poiché inizialmente il suo livello sarà troppo basso.
Attualmente, l'Editor della Community di Moovit è disponibile in tutte le sue funzionalità esclusivamente via web tramite un browser che riporta uno *user-agent* desktop.

Oltre all'editor delle fermate e delle linee, nell'ultimo periodo Moovit ha messo a disposizione direttamente dall'app mobile uno strumento che permette di segnalare anche l'affollamento di una determinata corsa, come riportato sul sito web della piattaforma.[4]

[4] `https://moovit.com/press-releases/moovit-crowding-report/`

2.3 Il focus di ShallWeGo

La piattaforma ShallWeGo si basa principalmente sul modello delineato da Moovit con la sua Community rendendo tuttavia i dati ottenuti direttamente dagli utenti la fonte principale di informazioni, fornendo gli strumenti adatti per rendere disponibili i dettagli sulle aziende, sulle corse e sulle fermate in una determinata zona tramite un'applicazione mobile che ne semplifichi l'accessibilità anche all'esterno in quanto, come accennato in precedenza, la piattaforma di Moovit risulta principalmente web-based, separata dall'applicazione principale che è più conosciuta al grande pubblico.
Inoltre, a differenza della Community di Moovit, in ShallWeGo le segnalazioni non sono limitate solamente a dettagli statici ma si estendono anche ad eventi "dinamici" (come strade chiuse, deviazioni, traffico o incidenti), ai dettagli di una singola fermata (come l'affollamento, la presenza di pensiline, di paline di riconoscimento delle aziende che la utilizzano o dei quadri orari) ed all'andamento delle corse, sfruttando il dispositivo dell'utente, permettendo quindi ad altri membri della community di verificare la presenza e la posizione di una corsa nell'ambito di una linea, in modo simile a quanto avviene con Transit Realtime di Google, ma con la differenza che non risulta necessario che le aziende condividano i dati. La feature presente in ShallWeGo permette anche di condividere informazioni sulla corsa stessa quali l'affollamento, la temperatura o le condizioni del mezzo in quel momento (come ad esempio lo stato di funzionamento delle obliteratrici o dell'aria condizionata)
Infine, dal punto di vista della verifica delle segnalazioni, ShallWeGo non possiede un team che si dedica a valutare le segnalazioni, ma al contrario i verificatori più adatti vengono scelti mediante una componente di Intelligenza Artificiale creata proprio a questo scopo ed integrata nell'infrastruttura dell'applicazione. In particolare, l'Algoritmo per la scelta degli utenti si basa, oltre che sull'esperienza accumulata sulla piattaforma da parte di un singolo (in termini di segnalazioni effettuate e valutate) anche sulla distanza geografica dell'area in cui egli opera dal luogo della segnalazione. In questo modo, anche i nuovi iscritti alla piattaforma avranno la possibilità di valutare le segnalazioni inviate alla piattaforma. I dettagli sul funzionamento della componente di Intelligenza Artificiale saranno discussi nel capitolo successivo.

CAPITOLO 3

Background sull'Intelligenza Artificiale utilizzata

Lo scopo di questo capitolo è di illustrare in dettaglio la componente di Intelligenza Artificiale che sta alla base del funzionamento della piattaforma.

3.1 Descrizione del problema da affrontare

Come descritto in precedenza, la fonte dei dati disponibili in ShallWeGo risiede nella sua community di utenti, che mettono a disposizione la loro conoscenza del *topic* con gli altri. Il concetto base su cui si basa la piattaforma è quindi la **segnalazione** da parte del singolo. Di questi dati, non essendo forniti direttamente dalle aziende di trasporto, non è però garantita la precisione o addirittura la correttezza. Sorge quindi il bisogno di effettuare una qualche tipo di **validazione**.
Le persone più adatte a validare i dati di una certa segnalazione sono quelle che possiedono principalmente i seguenti requisiti:

- Risiedano (oppure operino abitualmente) in una zona vicina al luogo oggetto della segnalazione.
- Abbiano una certa "reputazione" all'interno della piattaforma. (ShallWeGo infatti tiene traccia del livello di attività in termini di segnalazioni e di verifica delle stesse. Questo livello, similmente a quanto avviene sul social network Reddit, viene chiamato "*karma*" e cresce all'aumentare dell'attività del singolo)

I "verificatori" sono assegnati all'atto dell'invio di una segnalazione da parte dell'utente. Se ci si pone nello scenario in cui la piattaforma cresce in termini di numero di utenti, una ricerca esaustiva all'interno dell'insieme degli iscritti per assegnare quelli più adatti risulta essere problematica in termini di complessità. Si è quindi deciso di sfruttare una tecnica particolare che permetta di evitarla: quella degli **Algoritmi Genetici**, particolarmente utili alla ricerca in domini molto grandi.

3.2 Algoritmi genetici

Un algoritmo genetico non consiste tanto in una tipologia vera e propria di algoritmi quanto in un paradigma (una **metaeuristica**) che consente di implementare un algoritmo di **ricerca** per risolvere problemi di ottimizzazione. In quanto metaeuristica, il paradigma che sta alla base degli algoritmi generici non garantisce di trovare la soluzione ottima nello spazio di ricerca che si sta considerano ma permette di trovare in maniera molto veloce (e in questo risiede uno dei suoi vantaggi che lo fanno preferire ad altri approcci) delle soluzioni che vi si avvicinano. Un altro vantaggio che rende molto utile l'impiego di algoritmi che si basano su questa metaeuristica, consiste nel fatto che essi siano sempre applicabili, a prescindere dalla

struttura del problema.
In particolare, ad un algoritmo che sfrutta questa metaeuristica viene presentato un insieme iniziale di soluzioni ammissibili al problema che si sta affrontano e, secondo determinate metriche che forniscono una stima di quanto *"buona"* sia una certa soluzione a quel determinato problema, stabilire la migliore (o le migliori), andando a crearne di nuove se necessario.

3.3 Struttura e funzionamento degli algoritmi genetici

In generale, il funzionamento degli algoritmi genetici (e da qui la denominazione del paradigma) ricalca a grandi linee quella dell'evoluzione delle specie descritte da Charles Darwin nel 1854 nella sua opera *"L'origine della specie"* per cui l'evoluzione delle specie presenti in natura segue un cammino ben preciso, che consiste in tre fasi ben precise, a partire da una popolazione iniziale di individui:

- Selezione naturale
- Accoppiamento tra individui (detto anche, nel contesto che si sta trattando, *crossover*)
- Mutazione di un gene di un individuo

Un individuo, secondo quanto descritto da Darwin, può essere visto come un insieme di caratteristiche (o *geni*) che ne definiscono l'identità. Mediante l'accoppiamento i geni di due individui si mescolano, risultando nella creazione di un altro individuo che possiederà quindi una combinazione di quelli dei suoi genitori.
Con una certa probabilità, inoltre, avviene il fenomeno della mutazione, descritto poc'anzi.

Infine, sulla base di quanto buone siano le caratteristiche di un individuo, esso potrà "sopravvivere" ed eventualmente riprodursi per formare nuovi individui o "morire", non propagando oltre i suoi geni, che evidentemente non erano abbastanza adatti all'ambiente in cui si trovava. Il processo quindi risulta *iterativo*.

La strategia utilizzata da un algoritmo che implementa questo paradigma è particolarmente simile a quella descritta da Darwin per il suo campo di studi.

Portando avanti la similitudine con quanto avviene in natura, quando si vuole affrontare un dato problema (che chiameremo x) mediante la tecnica degli algoritmi genetici innanzitutto vengono generate in maniera più o meno casuale una serie di soluzioni ammissibili per x, composte da differenti caratteristiche (i *geni*). Questo insieme di soluzioni al problema rappresenta la cosiddetta **popolazione iniziale**. A partire da quest'ultima, viene solitamente

calcolato l'indice di "bontà" di ogni singola soluzione (che in gergo viene chiamata **fitness**), tramite una funzione apposita, chiamata appunto ***funzione di Fitness***. Una volta valutata questa quantità è necessario stabilire quanti e quali individui possano essere ammessi a riprodursi.
Per questa ragione, a questo stadio si affronta solitamente la fase di **selezione**, che consiste nel far sopravvivere solamente gli individui migliori in termini di fitness. Gli individui che sopravvivono alla selezione sono ammessi al cosiddetto **mating pool**, ovvero saranno membri dell'insieme degli individui che si possono riprodurre.
Una volta effettuato l'accoppiamento, la popolazione risultante è rappresentata da un nuovo insieme di individui che, come accennato in precedenza, posseggono una parte dei geni del primo genitore ed una parte dei geni del secondo, mutuandone quindi le loro caratteristiche.
Si passa poi all'ultimo stadio dell'evoluzione, in cui avviene (con una certa probabilità) la **mutazione** di uno o più geni che compongono un individuo all'interno della popolazione.
A questo punto l'algoritmo si trova davanti ad una scelta:

- Ricominciare dall'operazione di selezione usando come popolazione quella risultante dall'operazione di mutazione
- Terminare il processo

Per prendere una decisione, l'algoritmo fa riferimento alla cosiddetta **condizione di terminazione**, che rappresenta le condizioni che si devono verificare per mettere fine al processo (come ad esempio il numero di iterazioni effettuate, il tempo di esecuzione o il peggioramento in termini di fitness media della popolazione dopo un numero n fissato di iterazioni). Se la condizione di terminazione viene soddisfatta, allora il processo giunge al termine e solitamente viene restituito l'individuo migliore in termini di fitness dell'ultima popolazione, al netto di accorgimenti particolari e di operazioni di post-processing.

Le operazioni di selezione, crossover e mutazione vengono implementate tramite procedure chiamate **operatori genetici**, la cui definizione rappresenta uno degli step più importanti dell'implementazione di un algoritmo di ricerca che sfrutta questa metaeuristica, assieme a stabilire come debba essere calcolata la fitness e come e quando debba terminare il processo.

3.4 L'algoritmo utilizzato da ShallWeGo

In questa sezione verrà descritto accuratamente l'algoritmo di ricerca utilizzato dalla piattaforma che segue il paradigma degli algoritmi genetici e che va a risolvere il problema di

ottimizzazione consistente nel selezionare i migliori potenziali verificatori di una determinata segnalazione.

3.4.1 Codifica dell'Individuo

Nell'algoritmo presente in ShallWeGo, si definisce individuo un insieme di 5 utenti (nella popolazione iniziale essi sono presi in maniera casuale da quelli che operano nella provincia della segnalazione. In questo modo, sfruttando la vicinanza territoriale, viene aumentata la possibilità che questi utenti siano a conoscenza dell'oggetto della segnalazione per la quale sono stati scelti. La distanza tra il luogo della segnalazione e quello in cui opera l'utente viene calcolata effettuando delle operazioni di geocoding tramite il server Nominatim messo a disposizione per l'occasione. Questo aspetto verrà trattato in dettaglio nel capitolo dedicato all'architettura della piattaforma.

3.4.2 Funzione di Fitness

All'inizio di questo capitolo, è stato delineato l'*identikit* del verificatore ideale per una certa segnalazione. Le metriche utilizzate, riassumendo, sono:

- La distanza in chilometri tra il luogo di una segnalazione e l'area in cui opera un utente.
- La "reputazione" (o karma) che un utente si è costruito durante la sua permanenza sulla piattaforma

Dal punto di vista matematico, la funzione di fitness può essere formalizzata in questo modo, se x rappresenta il valore (naturalmente $>= 0$) in chilometri della distanza ed y il valore di karma dell'utente:

$$f(x,y) = \begin{cases} \frac{4(250+\frac{100}{x+0,3})+2y}{2} & \text{se } x < 15 \\ \frac{4(\frac{30}{x})^2+2y}{2} & \text{altrimenti} \end{cases}$$

La struttura dell'equazione conferma quanto accennato poc'anzi sul dare la priorità alla distanza geografica piuttosto che (almeno in un primo momento della vita della piattaforma, con una distribuzione più uniforme degli utenti) al karma. Infatti, un utente che opera in un'area che dista meno di **15km** dal luogo della segnalazione ha intuitivamente più probabilità di essere a conoscenza della struttura della rete di trasporti in quella zona (e quindi risulta più adatto al ruolo di verificatore).
Ed infatti, visto l'andamento della funzione $f(x) = \frac{100}{x}, x \neq 0$ (il cui andamento è mostrato

nella figura successiva e che a valori piccoli di x associa valori grandi di $f(x)$) e vista la presenza della costante additiva (250), un utente che opera in quel range di distanza dalla segnalazione risulta molto forte agli occhi dell'algoritmo.

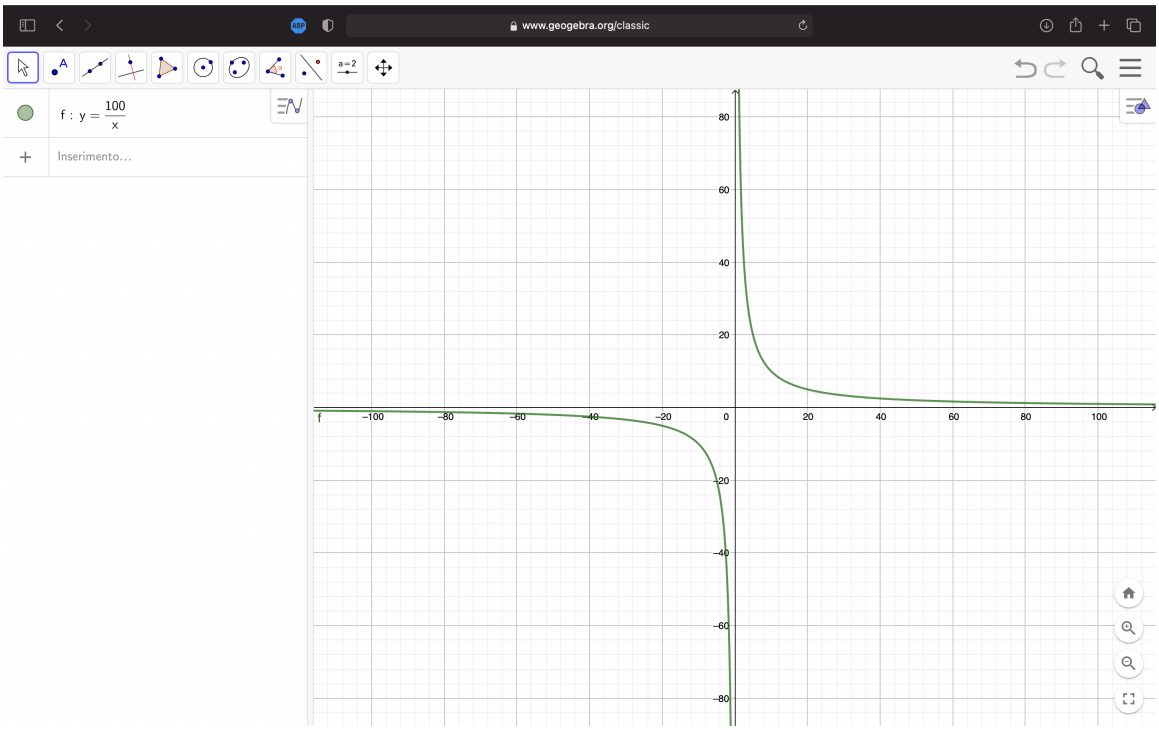

Figura 3.1: L'andamento della funzione $100/x$

3.5 Operatori genetici

Verranno ora illustrate le scelte riguardo all'implementazione dei vari operatori genetici che prendono parte al processo di evoluzione tipico degli algoritmi genetici.

3.5.1 L'operatore genetico di Selezione

Il primo operatore che viene applicato in una singola iterazione dell'algoritmo è quello di **Selezione**, che, riprendendo quanto descritto in precedenza, permette di stabilire a quali individui possano essere ammessi a riprodursi ed eventualmente a mutare (sulla base della loro fitness). In letteratura, esistono diversi approcci a quest'operatore. Quelli principali sono stati illustrati durante la sesta lezione del corso di *Fondamenti di Intelligenza Artificiale* tenuta dal Prof. Fabio Palomba e dal Dott. Emanuele Iannone.
Tra questi vanno menzionati in particolare:

- L'approccio basato su **roulette wheel** in cui gli individui con la fitness più alta hanno più possibilità di sopravvivere a questa fase, proprio come succede ad un elemento posto su una roulette che ha una porzione di area maggiore rispetto agli altri. In questo caso, la "porzione" di roulette dedicata ad un individuo è direttamente proporzionale al valore della sua fitness (normalizzata in $[0, 1]$)

- L'approccio basato su **rank**, nel quale ogni individuo della popolazione è ordinato in modo decrescente rispetto alla sua fitness e ad ogni posizione in questa lista viene associato un numero, il cosiddetto *rango*: la probabilità di selezione del singolo risulta inversamente proporzionale al suo rango.

- L'approccio basato su **truncation**, in qualche modo simile al precedente (ne mutua l'ordinamento degli individui) che consiste nel fissare un numero $n < |P|$ (se P rappresenta la popolazione) e di selezionare i primi n individui nella popolazione che hanno la fitness più alta

Per l'algoritmo utilizzato in ShallWeGo si è scelto di utilizzare l'approccio basato su **roulette wheel** in quanto risulta quello di più immediata comprensione ed implementazione (essendo molto più fedele a quanto avviene in natura). Inoltre, non essendoci alcun caso in cui la fitness possa scendere sotto il valore 0 quest'approccio è stato applicabile senza problemi.

3.5.2 L'operatore genetico di Crossover

Dopo l'applicazione dell'operatore di Selezione (e quindi con al definizione di quelli che sono i componenti del cosiddetto **mating pool**), si procede con la fase di **crossover**, che prevede l'accoppiamento (con una determinata probabilità) di coppie di individui all'interno del mating pool. Esistono diverse tecniche che permettono di implementare questa operazione. Tra queste si menzionano:

- La tecnica **Single Point** che dati due individui x = $(x_0, x_1, ..., x_n)$ ed y = $(y_0, y_1, ..., y_m)$, prevede la creazione di due nuovi individui j e k che siano il risultato di un incrocio tra i geni dei loro genitori, dopo aver selezionato il cosiddetto **punto di taglio**, ovvero quel punto z tale che:

  - $j = x_0, x_1, ..., x_z, y_{z+1}, ..., y_m$ e
  - $k = y_0, y_1, ..., y_z, x_{z+1}, ..., x_n$

- La tecnica ***k*-points**, che rappresenta una generalizzazione della tecnica Single Point, in cui un individuo viene diviso in k porzioni prima di essere incrociato con un altro individuo

La tecnica che viene usata da ShallWeGo è quella del **Single Point**. C'è da precisare che nel caso del problema che si sta affrontando la posizione di un gene all'interno di un individuo risulta irrilevante e di conseguenza un approccio k-points non avrebbe portato ad un qualche tipo di miglioramento.

3.5.3 L'operatore genetico di Mutazione

L'ultimo operatore genetico che viene applicato durante un'iterazione è quello di **Mutazione** che ricalca, come accennato precedentemente, il processo di mutazione spontanea che avviene in natura. Una mutazione consiste nel cambiamento casuale di un gene all'interno di un individuo. Anche in questo caso, esistono diversi approcci che in generale possono essere seguiti.
Tra questi si menzionano:

- La tecnica del **Bit Flip** che, come suggerisce il nome, è applicabile solo nel caso di individui codificati in modo *binario* e consiste nel scegliere in maniera casuale un gene all'interno dell'individuo e "capovolgere" il suo valore (quindi da 0 si passa ad 1 e vice-versa).

- La tecnica del **Random Resetting** che prevede il cambiamento di un gene preso casualmente all'interno dell'individuo con un altro valore ammissibile per quel gene preso altrettanto casualmente.

- La tecnica dello **Swap** che prevede lo scambio di due geni all'interno dell'individuo

- La tecnica dello **Scramble** che prevede una permutazione dei geni all'interno di un individuo

La maggior parte di questi approcci risultano piuttosto rilevanti in un contesto in cui la posizione di un gene cambia la sua rilevanza del suo valore in un individuo, come nel caso di codifiche binarie (il bit più a sinistra se è posto ad 1 aumenta di gran lunga il valore del numero rappresentato in binario).
Data la particolare natura del problema che si sta affrontando, non è possibile andare ad utilizzare la tecnica del Bit Flip e, nello stesso modo, le tecniche di Swap e Scramble non

sono rilevanti in quanto la codifica dell'individuo non risente della posizione di un gene in particolare.
Di conseguenza, la strategia che viene usata per questo algoritmo è quella del **Random Resetting**. In particolare, scelto un utente che è presente all'interno di un individuo, lo si sostituisce con un nuovo utente preso dal pool dei potenziali candidati ad essere verificatori di una segnalazione. Nel caso il nuovo utente che ha sostituito il precedente fosse già presente all'interno dell'individuo, il processo si ripete fino a non avere utenti uguali.

3.6 Termine del processo ed accorgimenti adottati

3.6.1 Condizione di terminazione

L'algoritmo termina quando si verifica una delle seguenti tre condizioni:

- Vengono superate le 25 iterazioni del processo
- Viene superato il limite di tempo stabilito (nel caso dell'algoritmo della piattaforma, quest'ultimo è fissato a 5 minuti)
- Per tre iterazioni consecutive la fitness della popolazione è minore rispetto a quella della miglior popolazione creata fino a quel momento

3.6.2 La strategia dell'archivio

Come ulteriore accorgimento nello sviluppo dell'algoritmo per la piattaforma, è stata implementata la cosiddetta **Strategia dell'Archivio** che consiste nel conservare una popolazione separata che non evolve composta da individui particolarmente forti, che si va a costruire man mano che le iterazioni vanno avanti. Nel caso specifico di ShallWeGo si è provveduto, per ogni iterazione, a tenere traccia del miglior individuo in termini di fitness della popolazione risultante dall'applicazione dei tre operatori genetici. Il migliore tra questi tre verrà quindi aggiunto all'archivio. Essendo costituito da individui "forti", l'archivio può essere usato come alternativa alla popolazione che viene restituita dopo il termine della computazione dell'algoritmo. Questo, infatti, è proprio l'approccio scelto per la piattaforma: se al termine delle iterazioni la popolazione risultante avrà una fitness media minore o uguale a quella dell'archivio, la popolazione restituita sarà sostituita dall'archivio.
La popolazione restituita sarà quindi soggetta a delle operazioni di postprocessing.

3.6.3 Postprocessing

Dopo l'esecuzione dell'algoritmo il risultato è quindi una popolazione di individui, a loro volta composta da utenti. In precedenza, è stato specificato come la fitness di un individuo sia calcolata tramite una media aritmetica della fitness calcolata sui singoli geni. Si sfrutta quindi questa possibilità per effettuare un lavoro di post-processing sulla popolazione risultato. In particolare, si è scelto di effettuare le seguenti operazioni:

- Si ottiene, a partire dal singolo individuo, la lista degli utenti che esso contiene, ordinati in maniera decrescente rispetto al loro valore di fitness;
- Per ognuna di queste liste, viene preso il primo elemento. Il risultato di questa operazione sarà di fatto un nuovo individuo formato dagli utenti più "forti" di quelli presenti tra gli individui della popolazione risultante dall'algoritmo.

Sebbene le operazioni di post processing (trattandosi prettamente di ordinamento) assumano una complessità di almeno $\mathcal{O}(n \log n)$, esse vengono effettuate su un dominio particolarmente più piccolo rispetto a quello dei candidati (che mettendosi nel contesto di una piattaforma ben avviata potrebbero essere in numero molto elevato o comunque tale da rendere la ricerca esaustiva poco efficiente) e che allo stesso tempo permettono di ottenere un risultato ancora migliore rispetto a quello ottenuto con la semplice evoluzione, riuscendo cioè a creare un gruppo di utenti che risultino "forti tra i forti" in termini di fitness assoluta. Ciò giustifica la presenza sia della procedura di evoluzione degli individui sia il postprocessing. Le operazioni di ordinamento non richiedono un ulteriore calcolo della fitness in quanto in fase di implementazione è stato effettuato il caching di quel valore per quella determinata istanza dell'algoritmo e sicuramente questo valore alla fine sarà già stato calcolato e conservato, riducendo il calcolo finale della fitness ad un semplice accesso ad una variabile che non richiede operazioni particolari.

3.6.4 Risultati dell'algoritmo con i parametri correnti

Nella sua fase di testing, l'algoritmo è stato eseguito su un insieme molto grande di utenti generato in maniera casuale (un utente per ogni comune e con fitness presa casualmente tra 0 e 65). Nella tabella che segue sono riportati i risultati dell'algoritmo su diverse istanze. In particolare,

- Nella **prima colonna** è riportato il comune della segnalazione (e le sue coordinate);

- Nella **seconda colonna** sono riportati comune di residenza e livello di karma degli utenti selezionati dall'algoritmo per verificare quella determinata segnalazione, dopo l'applicazione del post-processing;

Fisciano	Cava de' Tirreni (karma: 48.6) Siano (karma: 12.5) Baronissi (karma: 17.8) Roccapiemonte (karma: 47.8) Trentinara (karma: 53.6)
Nocera Inferiore	Nocera Inferiore (karma: 29.2) Sant'Egidio del Monte Albino (karma: 39.3) San Marzano sul Sarno (karma: 53.4) Roccapiemonte (karma: 47.8) Nocera Superiore (karma: 24.2)
Salerno	Giffoni Sei Casali (karma: 45.3) Pontecagnano Faiano (karma: 44.2) Baronissi (karma: 17.8) Laviano (karma: 46.8) Maiori (karma: 46.6)

Tabella 3.1: Risultati dell'algoritmo su diversi input.

Si noti come nel caso dell'istanza dell'algoritmo per la città di Nocera Inferiore sia presente un utente che opera proprio lì: egli rappresenta quindi parte della soluzione ottima. Nel caso della città di Salerno, invece, è stato incluso un utente la cui distanza dal luogo della segnalazione si aggira attorno ai 47km in linea d'aria. Quest'utente è stato incluso nella soluzione in virtù del proprio valore di *karma* che risulta discretamente alto.

CAPITOLO 4

ShallWeGo

Lo scopo di questo capitolo è di descrivere in termini di implementazione il funzionamento e l'architettura della piattaforma.

4.1 Architettura del sistema

Il sistema ShallWeGo si configura come una classica architettura **client-server**.

- Il server è rappresentato da una *Java Enterprise application*
- Il client è stato realizzato tramite un'applicazione utilizzabile sulla piattaforma Android dalla versione 10 in poi, a causa di scelte implementative riguardo alcune librerie incluse nel progetto. Queste scelte saranno documentate più avanti nel capitolo.

4.2 Implementazione: framework e librerie utilizzate

Il server, come precedentemente menzionato, è realizzato usando le specifiche Java Enterprise Edition (da qui in poi *Java EE*), usando a questo scopo il framework ***Spring Boot***, che permette una più semplice realizzazione dei task necessari a sviluppare un'applicazione che sfrutta il paradigma Client-Server, come ad esempio la gestione delle API accessibili dall'esterno i cosiddetti *endpoint* oppure la gestione dei dati persistenti.

4.2.1 Gestione dei dati persistenti: breve introduzione a JPA

Per quanto riguarda quest'ultimo aspetto, Spring supporta il cosiddetto *Object-Relational Mapping* (**ORM**).

ORM è definito come "una tecnica di programmazione che permette l'integrazione di sistemi software che aderiscono al paradigma della programmazione orientata agli oggetti (**OOP**) con sistemi **RDBMS** (*Relational DataBase Management System*)".

Uno dei principali vantaggi di ORM risiede nel contrasto della complessità di gestione della persistenza derivata dalla mancanza di compatibilità tra dati salvati all'interno di un database e oggetti in un linguaggio di programmazione, oltre che ad una sostanziale indipendenza dal *vendor* che fornisce il DBMS utilizzato. Wikipedia [2021]

Spring, nello specifico, mette a disposizione una serie di API che implementano le specifiche di **JPA** (Java Persistence API). JPA è definito come "un insieme di specifiche che mette a disposizione degli sviluppatori un insieme di servizi e di strutture dati che permettono l'ORM e quindi la gestione dei dati persistenti all'interno di applicazioni Java." (The Java EE 7 Authors [2014]). Essendo solamente una specifica, JPA necessita di un'implementazione. Quella di riferimento è **EclipseLink**, sviluppata dalla Eclipse Foundation a partire dal 2015.

Il framework Spring, tuttavia, utilizza come implementazione delle specifiche JPA la libreria open source **Hibernate 5.5** sviluppata in partnership con l'azienda *Red Hat*.

Per effettuare il mapping tra oggetti istanze di una classe e righe presenti in una tabella in un database, JPA si serve del concetto di **Entity**, che rappresenta un'istanza di una classe (in gergo, **POJO**, ovvero *Plain Old Java Object*) che può essere salvata all'interno di un database.

Le specifiche JPA introducono il concetto di **Configuration By Excpetion**. In questo modo, se non specificato altrimenti, le impostazioni riguardo persistenza e mapping di oggetti risulteranno essere quelle di default. Per fare un esempio concreto, in un progetto dove è previsto l'utilizzo di JPA tutte le classi Java vengono viste come tali fino a che non viene utilizzata l'annotazione Entity su una di queste. Ad esempio:

```
@Entity
public class POJO implements Serializable {
    @Id
    private Integer id;

    public POJO() {}
}
```

Listing 1: File: POJO.java

Tramite questo setup è possibile rendere la classe contenuta in *POJO.java* un'entity di cui è possibile effettuare la persistenza su un database.

Per poter rappresentare un'Entity, un POJO deve rispettare i seguenti requisiti: (Goncalves [2013])

- Essere annotata con @javax.persistence.Entity
- Avere un attributo annotato con @javax.persistence.Id che ne denota la chiave primaria (nel caso di chiave primaria semplice: è possibile anche avere chiavi composte, come descritto più avanti)
- Deve possedere un costruttore vuoto che abbia come modificatore di accesso *public* o *protected*. Sono ammessi ulteriori costruttori.
- Non deve essere una classe *final*
- Non deve essere una classe interna ad un'altra classe
- Deve implementare l'interfaccia `Serializable`.

Le Entity presenti nel dominio applicativo sono le seguenti:

- **Stop**, che modella una fermata dei mezzi pubblici,
- **Line**, che modella una linea di trasporto pubblico,
- **Company**, che modella un'azienda di trasporto pubblico,
- **Report**, che modella il concetto di segnalazione,
- **LineReport**, che modella il concetto di segnalazione di una linea,
- **CompanyReport**, che modella il concetto di segnalazione di un'azienda,
- **StopReport**, che modella il concetto di segnalazione di una fermata,
- **TemporaryEventReport**, che modella il concetto di segnalazione di un evento temporaneo,
- **User**, che modella un utente registrato alla piattaforma.

oltre che ad altre classi "di servizio" create per mappare relazioni di tipo Many-to-Many.

Si ponga attenzione sulla modellazione del concetto di segnalazione. Qui di seguito è mostrato lo scheletro delle classi che modellano i vari tipi di segnalazione:

```
@Entity
public abstract class Report {
    @Id
    private Integer Id;
}
```

Listing 2: File: Report.java

```
@Entity
public class StopReport extends Report {
    private Stop stopReported;
}
```

Listing 3: File: StopReport.java

```
@Entity
public class LineReport extends Report {
    private Line lineReported;
}
```

Listing 4: File: LineReport.java

```
@Entity
public class CompanyReport extends Report {
    private Company companyReported;
}
```

Listing 5: File: CompanyReport.java

```
@Entity
public class TemporaryEventReport extends Report {
    private Date validityStart;
    private Date validityEnd;
    private String eventType;
    private String description;
    private String latitude;
    private String longitude;
    private String source;

    private List<Line> affectedLines;
}
```

Listing 6: File: TemporaryEventReport.java

Si noti l'utilizzo della keyword **extends** che segnala l'utilizzo dell'ereditarietà, un concetto tipico dei linguaggi Object-Oriented come Java e che non è modellato in nessun DBMS

relazionale. Per ovviare a questo problema, JPA mette a disposizione la seguente annotazione:

```
public @interface Inheritance {
    InheritanceType strategy() default SINGLE_TABLE;
}
```

Quest'annotazione, attraverso il campo *strategy* permette di stabilire con che strategia effettuare il mapping di una gerarchia.

I valori che *strategy* può assumere sono i seguenti: (Goncalves [2013])

- SINGLE_TABLE, che crea una sola tabella all'interno del database che rappresenta l'intera gerarchia e che per disambiguare tra i vari tipi delle sottoclassi usa un attributo chiamato "DTYPE" che assume come valore il nome della sottoclasse di un determinato elemento della tabella. Questa strategia, in virtù della configuration by exception è quella utilizzata di default in assenza dell'annotazione.
- TABLE_PER_CLASS, che crea una tabella all'interno del database per ogni sottoclasse in una strategia.
- JOINED, che colloca gli attributi che compaiono esclusivamente nelle sottoclassi della gerarchia in altre tabelle, una per membro della gerarchia.

In ShallWeGo, la strategia utilizzata è quella SINGLE_TABLE poiché si è preferito non aumentare la complessità dello schema.

CAPITOLO 5

Sviluppi futuri

CAPITOLO 6

Conclusioni

BREVE SPIEGAZIONE CONTENUTO CAPITOLO

Bibliografia

[Adnkronos 2019] ADNKRONOS: Raggi: Orari in tempo reale su GMaps. (2019). – URL https://www.adnkronos.com/raggi-orari-bus-in-tempo-reale-su-gmaps_3KnBSCR6JLTAX1ZwI5oLhS (Citato a pagina 8)

[Asstra 2021] ASSTRA, Agens: Trasporto pubblico locale – Interventi prioritari. (2021), S. 12 (Citato a pagina 1)

[De Girolamo 2021] DE GIROLAMO, Alfredo: TPL in ginocchio: due miliardi in meno di ricavi. (2021). – URL https://www.ilsole24ore.com/art/tpl-ginocchio-due-miliardi-meno-ricavi-AD4mLaFB (Citato a pagina 1)

[Goncalves 2013] GONCALVES, Antonio: *Beginning Java EE 7*. 1st. USA : Apress, 2013. – ISBN 143024626X (Citato alle pagine 24 e 27)

[The Java EE 7 Authors 2014] THE JAVA EE 7 AUTHORS: *The Java EE 7 Tutorial*. 7.0. https://docs.oracle.com/javaee/7/tutorial/persistence-intro.htm: , 2014 (Citato a pagina 23)

[Wikipedia 2021] WIKIPEDIA: *Object-Relational Mapping — Wikipedia, L'enciclopedia libera*. 2021. – URL https://it.wikipedia.org/wiki/Object-relational_mapping. – [Online; controllata il 19-agosto-2021] (Citato a pagina 23)

Siti Web consultati

- OpenStreetMap Nominatim – `https://nominatim.org`

Ringraziamenti

INSERIRE RINGRAZIAMENTI QUI