

Progetto piattaforma per ricerca Ordinamento Corso di Laurea

(Teodoro Montanaro)

Matricola 188924

(Ottobre 2012)

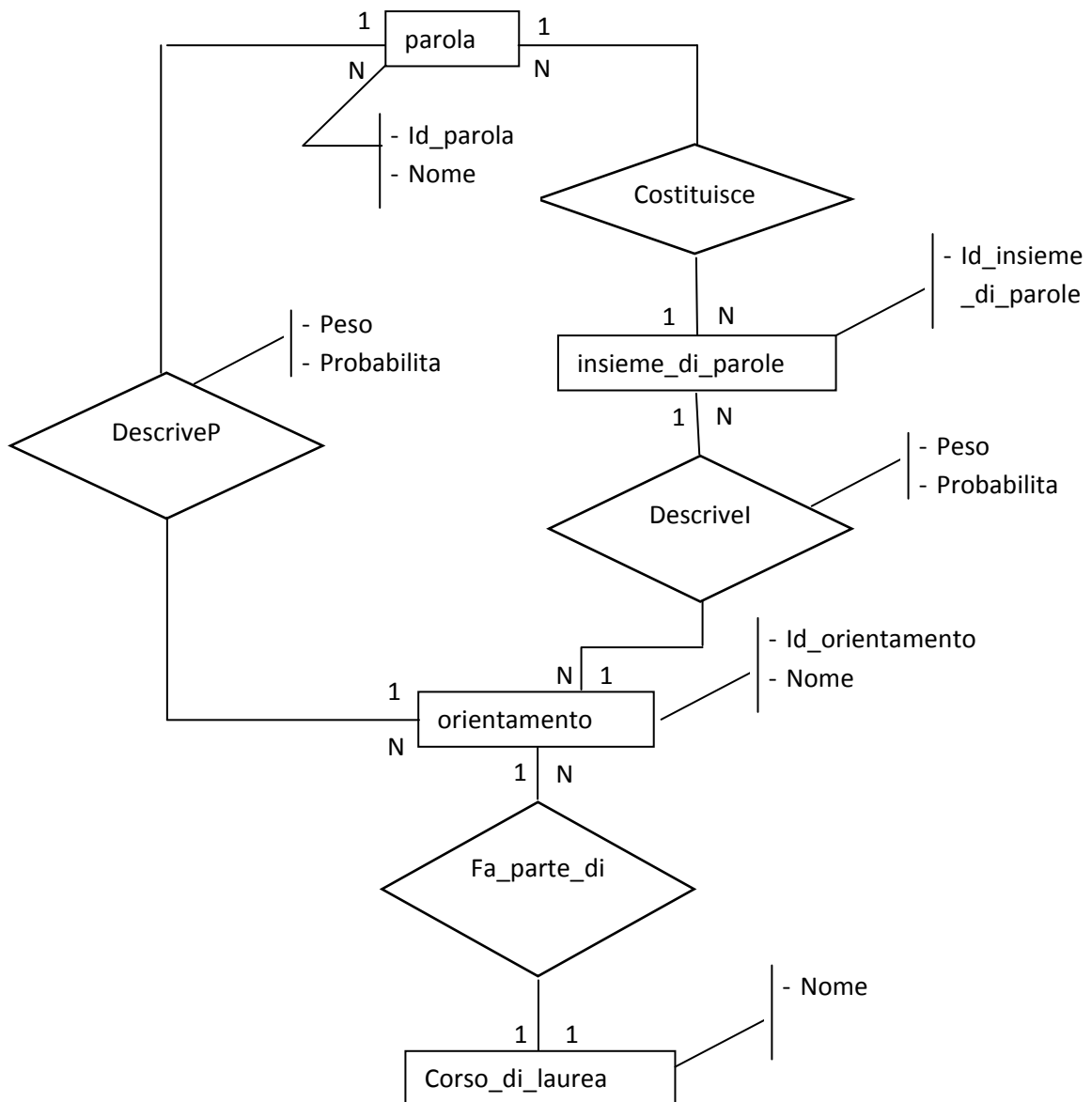
Sommario

1)	Obiettivo generale:.....	2
2)	Modellazione ER del database	2
3)	Dizionario dei dati.....	3
4)	Implementazione:.....	4
5)	Flow Chart algoritmo realizzato	6
6)	Come vengono calcolati i sinonimi	10
7)	Casi di studio.....	10
8)	Motivi per cui ha funzionato solo 2 volte su 4:	14
9)	Problemi irrisolti	14
10)	Possibili sviluppi futuri.....	15

1) Obiettivo generale:

Progettazione di una piattaforma che restituisca un elenco ordinato dei Corsi di Laurea (con relativo Orientamento) che più possono interessare la persona che effettua la ricerca. La ricerca è effettuata sulla base delle parole/frasi inserite dall'utente come descrizione dei propri interessi.

2) Modellazione ER del database



3) Dizionario dei dati

- *ENTITA' parola*: contiene tutte le parole caratteristiche per uno o più corsi/orientamenti
 - *id_parola*: rappresenta un valore progressivo assegnato a ciascuna nuova parola;
 - *Nome*: Contiene la parola vera e propria (è stata dichiarata **UNIQUE** in modo da non inserire più volte la stessa parola)
- *ENTITA' insieme_di_parole*: permette di associare più parole tra loro
 - *id_insieme_di_parole*: rappresenta un valore progressivo assegnato a ciascun nuovo insieme di parole
- *ENTITA' orientamento*: contiene tutti gli orientamenti di ogni corso di laurea (anche se ce n'è uno unico)
 - *id_orientamento*: rappresenta un valore progressivo assegnato a ciascun nuovo orientamento
 - *Nome*: indica il nome dell'orientamento
 - *Corso di laurea*: visto che la relazione tra orientamento e corso di laurea è di tipo 1 a molti si è pensato di non creare un'entità per ogni singolo corso di laurea, ma di inserire un campo nell'entità **ORIENTAMENTO** che contenga il corso
- *RELAZIONE costituisce*: mette in relazione più parole tra loro (passando attraverso l'entità insieme di parole)
- *RELAZIONE descrivep*: mette in relazione una singola parola con un orientamento
 - *peso*: indica che relazione c'è tra la parola e l'orientamento. Peso è un valore decimale che può essere compreso tra 0 e 1 con 5 posizioni dopo la virgola
 - Probabilità calcolata in base al peso usando la formula:

$$P(\text{parola} / \text{orientamento}) = \frac{\text{Peso}(\text{parola}, \text{orientamento})}{\sum \text{pesi di tutte le parole e insiemi di parole associati a orientamento}}$$

- *RELAZIONE descrivei*: mette in relazione un insieme di parole con un orientamento
 - *peso*: indica che relazione c'è tra l'insieme di parole e l'orientamento. Peso è un valore decimale che può essere compreso tra 0 e 1 con 5 posizioni dopo la virgola
 - Probabilità calcolata in base al peso usando la formula:

$$P(\text{insieme_parole} / \text{orientamento}) = \frac{\text{Peso}(\text{insieme_parole}, \text{orientamento})}{\sum \text{pesi di tutte le parole e insiemi di parole associati a orientamento}}$$

4) Implementazione:

- a) Evidenziazione delle parole e frasi più importanti nelle descrizioni dei vari corsi di Laurea
- b) Assegnazione di un peso ad ogni parola (valore compreso tra 0.00000 e 1.00000)

Il peso è stato assegnato sia all'associazione tra la parola e l'orientamento da cui è stata presa, sia alle associazioni con gli altri orientamenti che avevano in qualche modo a che fare con quella parola.

- c) Implementazione del database
- d) Creazione di uno script per il riempimento del database

Lo script prevede una minimale interfaccia tramite la quale è possibile inserire fino a 6 parole contemporaneamente.

Le parole inserite contemporaneamente verranno considerate come insieme di parole!

E' sconsigliato inserire articoli, preposizioni o qualsiasi cosa che può stare in mezzo tra 2 parole (se ad esempio devo inserire "Mezzi di comunicazione" inserirò solo le parole "Mezzi" e "Comunicazione")

inserisci.php

Parola1*:

Parola2:

Parola3:

Parola4:

Parola5:

Parola6:

Pesi (Numeri compresi tra 0.00000 e 1.00000) (Se lasci vuoto verrà considerato 0):

1 - Unico triennale - INGEGNERIA DEL CINEMA E DEI MEZZI DI COMUNICAZIONE

Peso1:

2 - Unico Triennale - INGEGNERIA DELLE TELECOMUNICAZIONI

Peso2:

3 - Unico Triennale - INGEGNERIA ELETTRONICA

Peso3:

4 - Unico triennale - INGEGNERIA FISICA

Peso4:

5 - Unico triennale - INGEGNERIA INFORMATICA

Peso5:

6 - Magistrale: Unico - FISICA DEI SISTEMI COMPLESSI (PHYSICS OF COMPLEX SYSTEMS)

- e) Realizzazione di uno script (calcola_prob.php) che una volta richiamato calcola la probabilità di ogni relazione tra parola e orientamento e tra insiemi di parole e orientamento secondo le formule:

$$P(\text{parola} / \text{orientamento}) = \frac{\text{Peso}(\text{parola}, \text{orientamento})}{\sum \text{pesi di tutte le parole e insiemi di parole associati a orientamento}}$$

$$P(\text{insieme_parole} / \text{orientamento}) = \frac{\text{Peso}(\text{insieme_parole}, \text{orientamento})}{\sum \text{pesi di tutte le parole e insiemi di parole associati a orientamento}}$$

- f) Implementazione dell' algoritmo che calcola l' elenco ordinato degli orientamenti che più si addicono alla persona che effettua la ricerca

La probabilità che un testo inserito sia corrispondente ad un particolare orientamento viene calcolata utilizzando le formule di Bayes:

$$P(\text{orientamentoA} / \text{testo}) = \frac{P(\text{orientamentoA}) * P(\text{parola1/orientamentoA}) * \dots * P(\text{parolan/orientamentoA})}{[P(\text{orA}) * P(\text{par1/orA}) * \dots * P(\text{parn/orA})] + \dots + [P(\text{orZ}) * P(\text{par1/orZ}) * \dots * P(\text{parn/orZ})]}$$

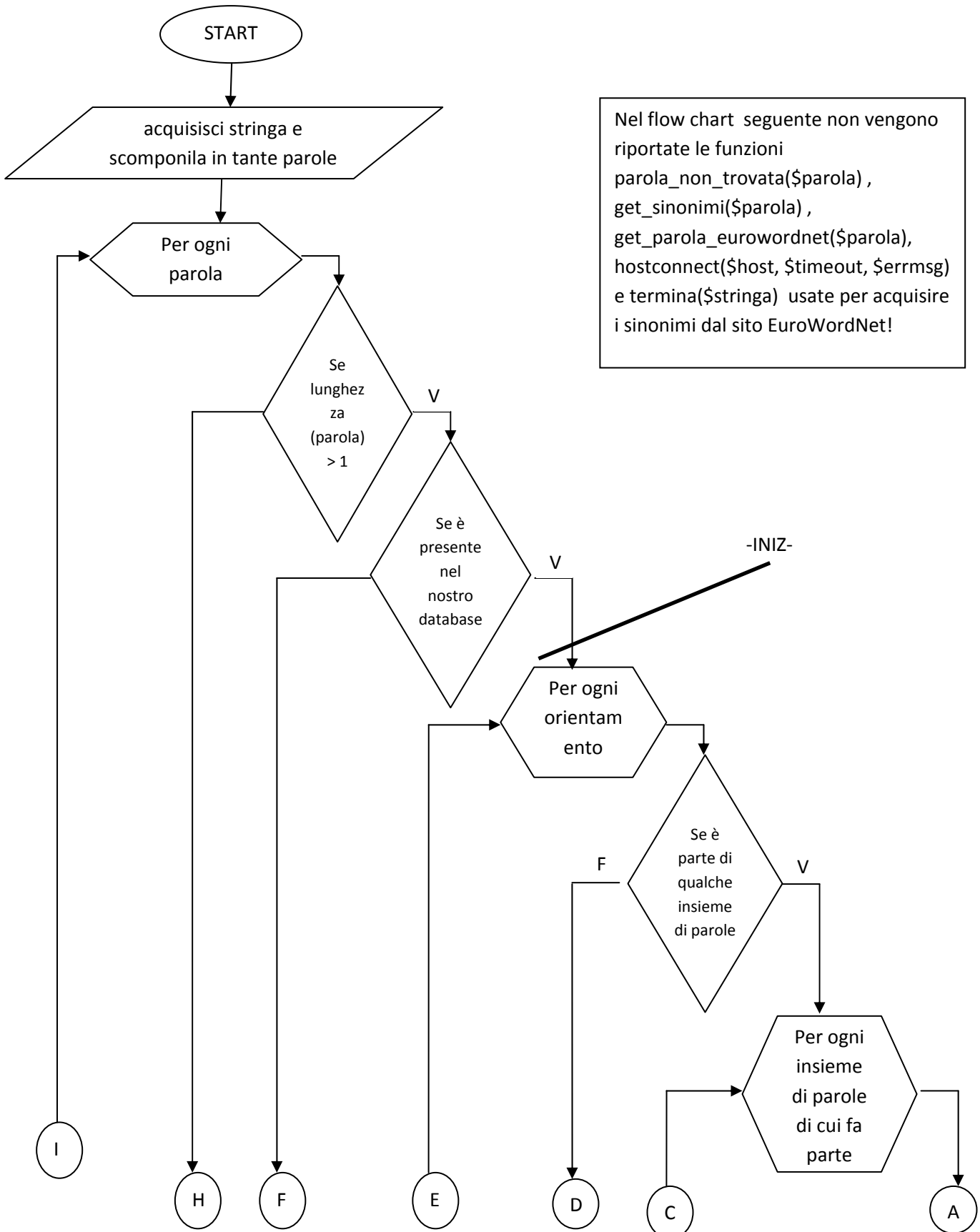
Dal momento che si suppone una distribuzione uniforme tutti gli orientamenti hanno la stessa probabilità di essere scelti, quindi si ha

$$P(\text{orientamentoA}) = P(\text{orientamentoB}) = \dots = P(\text{orientamentoZ})$$

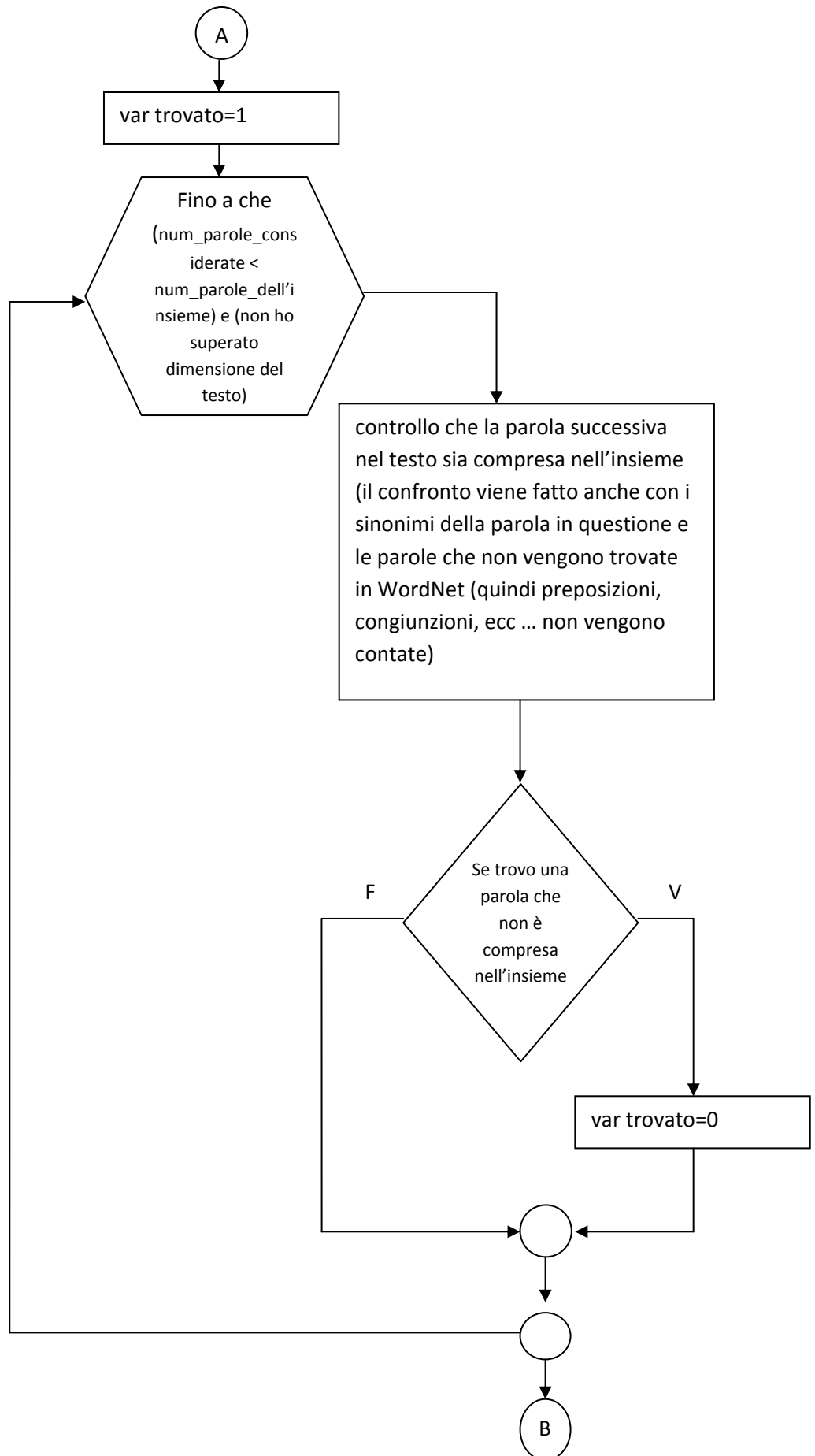
E di conseguenza si ha

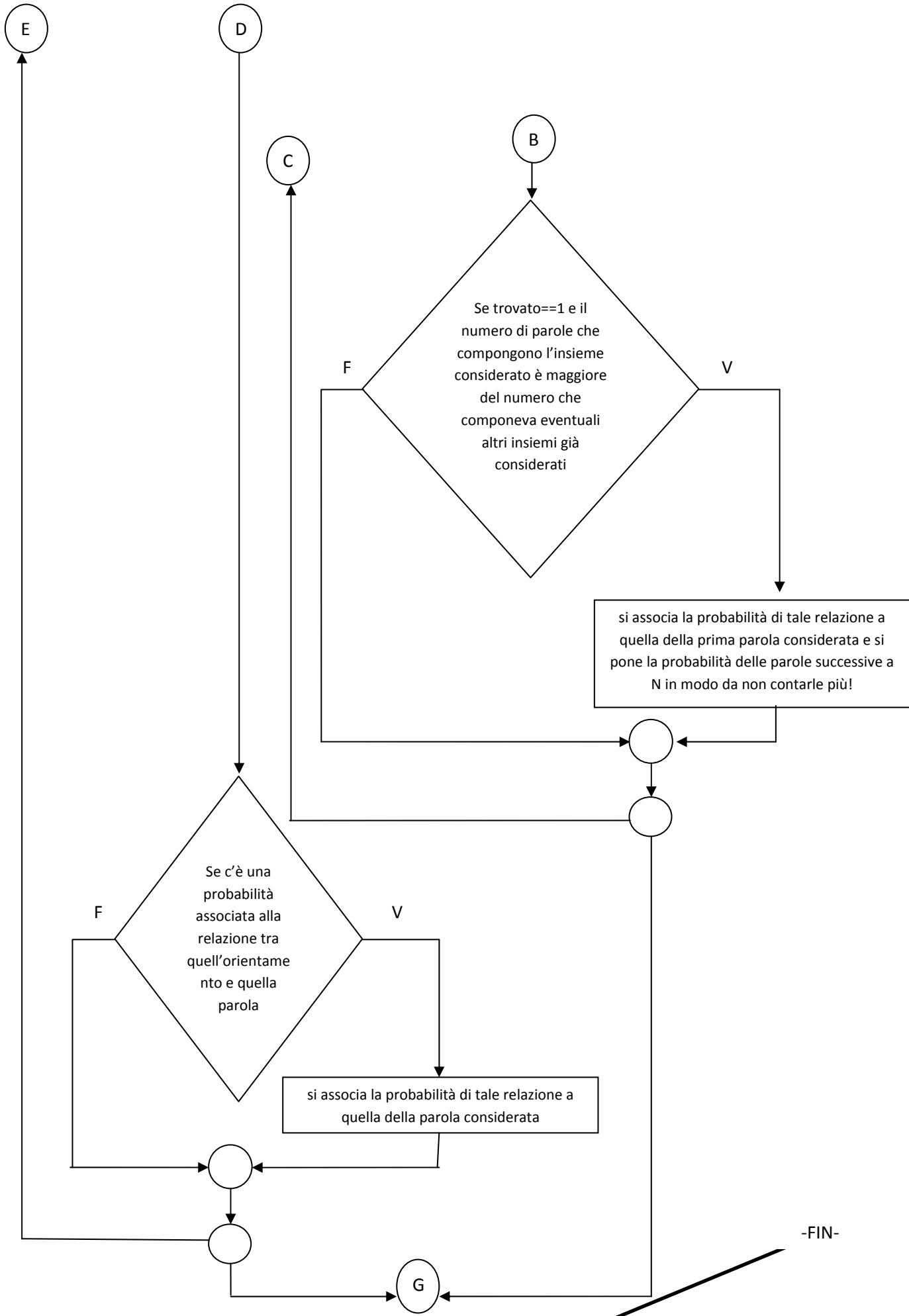
$$P(\text{orientamentoA} / \text{testo}) = \frac{P(\text{orientamentoA}) * P(\text{parola1/orientamentoA}) * \dots * P(\text{parolan/orientamentoA})}{[P(\text{orA}) * P(\text{par1/orA}) * \dots * P(\text{parn/orA})] + \dots + [P(\text{orZ}) * P(\text{par1/orZ}) * \dots * P(\text{parn/orZ})]} = \frac{\frac{1}{m} * P(\text{parola1/orientamentoA}) * \dots * P(\text{parolan/orientamentoA})}{\frac{1}{m} * \{[P(\text{par1/orA}) * \dots * P(\text{parn/orA})] + \dots + [P(\text{par1/orZ}) * \dots * P(\text{parn/orZ})]\}} = \frac{P(\text{parola1/orientamentoA}) * \dots * P(\text{parolan/orientamentoA})}{\{[P(\text{par1/orA}) * \dots * P(\text{parn/orA})] + \dots + [P(\text{par1/orZ}) * \dots * P(\text{parn/orZ})]\}}$$

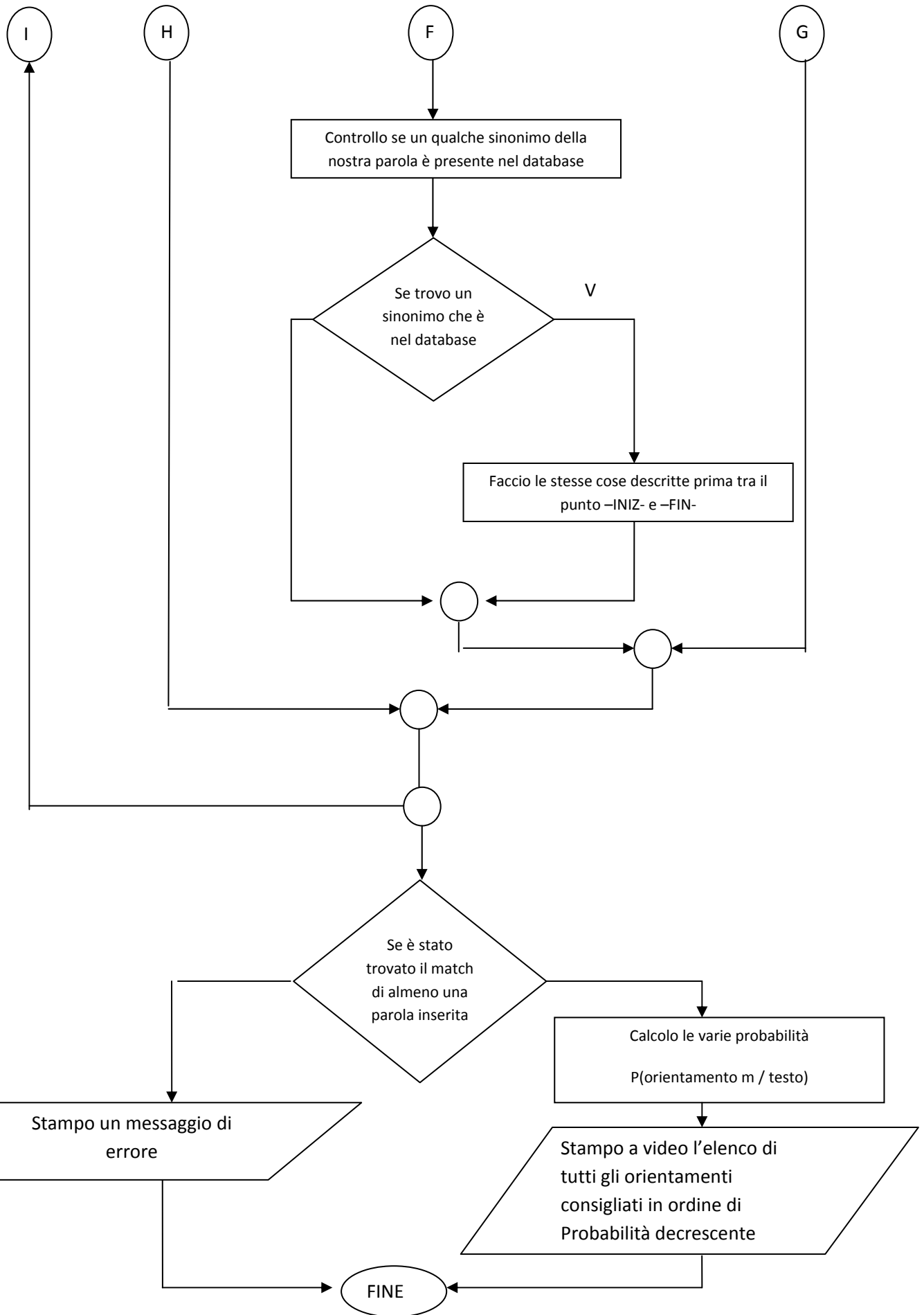
5) Flow Chart algoritmo realizzato



Progetto Intelligenza artificiale sviluppato da Teodoro Montanaro (matricola 188924)







6) Come vengono calcolati i sinonimi

Al momento, visto che non siamo in possesso del database di EuroWordNet e MultiWordNet, siamo costretti a collegarci al sito EuroWordNet (http://www.ilc.cnr.it/iwndb/iwndb_php/) e simulare una ricerca!

Nella pagina risultante otterremo tutti i sinonimi racchiusi tra i tag <i> e </i>!

Inoltre, EuroWordNet restituisce la scritta “parola non trovata” anche per gli articoli, le preposizioni ... quindi sfruttiamo tale funzione per verificare che una parola sia da cercare o meno!

7) Casi di studio

Sono stati presi 4 curriculum presenti sul web di persone laureate al Politecnico di Torino per vedere se l’algoritmo era in grado di scegliere proprio il percorso da loro intrapreso.

A parte il fatto che si è dovuta ridurre la mole di carico andando ad eliminare alcune parole proprio inutili (perché l’algoritmo si collega al sito EuroWordNet per acquisire i sinonimi delle parole),

Vi riportiamo i risultati:

1. http://www.istruzione.it/web/ministero/ministro/ministro_cv

Francesco Profumo -> Ingegnere Elettrotecnico

(parole cercate

I suoi interessi scientifici riguardano la conversione dell’energia, i componenti elettronici di potenza, sistemi innovativi nel campo dell’elettronica di potenza, sistemi integrati elettronici/elettromeccanici, azionamenti elettrici ad alta prestazione, nuove strutture di sistemi elettronici, sistemi per il condizionamento della potenza con utilizzo di celle a combustibile.

FIEEE-IAS

Conferenza IPEC

)

INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Automazione -> 19.31493 %
INGEGNERIA TELEMATICA (COMPUTER AND COMMUNICATION NETWORKS ENGINEERING) Magistrale: Unico -> 15.83727 %
INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Electronic micro and nanosystems -> 12.16251 %
INGEGNERIA DEL CINEMA E DEI MEZZI DI COMUNICAZIONE Unico triennale -> 12.10814 %
INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Embedded systems -> 11.11633 %
INGEGNERIA DEL CINEMA E DEI MEZZI DI COMUNICAZIONE Magistrale: Eurecom, Web Engineering, Multimedia -> 9.18457 %
INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Devices and Technologies for Integrated Electronics and Optoelectronics -> 6.46478 %
INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Sistemi elettronici -> 4.96860 %
INGEGNERIA ELETTRONICA Unico Triennale -> 3.79443 %
INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Progettazione analogica e di potenza -> 2.36038 %
INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Wireless Systems Design -> 0.56805 %
INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Communication Systems Engineering, Eurecom -> 0.47696 %
INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Embedded systems -> 0.45312 %
INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Software e sistemi digitali -> 0.33471 %
INGEGNERIA FISICA Unico triennale -> 0.28139 %
INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Fiber optic systems -> 0.19093 %

Progetto Intelligenza artificiale sviluppato da Teodoro Montanaro (matricola 188924)

INGEGNERIA DELLE TELECOMUNICAZIONI Unico Triennale -> 0.11893 %
 INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Information theory and signal processing applications -> 0.08375 %
 INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Wireless systems -> 0.07096 %
 FISICA DEI SISTEMI COMPLESSI (PHYSICS OF COMPLEX SYSTEMS) Magistrale: Unico -> 0.05092 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Software -> 0.01953 %
 INGEGNERIA MECCATRONICA (MECHATRONIC ENGINEERING) Magistrale: Unico -> 0.01768 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Reti -> 0.00768 %
 INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Multimedia e reti -> 0.00680 %
 INGEGNERIA INFORMATICA Unico triennale -> 0.00649 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Progettazione a RF -> 0.00005 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Grafica e computer animation -> 0.00004 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Multimedia -> 0.00004 %
 NANOTECNOLOGIE PER LE ICT (NANOTECHNOLOGIES FOR ICTs) Magistrale: Unico -> 0.00002 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Microelectronics -> 0.00000 %

2. <http://www.delet.polito.it/personale/scheda/%28nominativo%29/filippo.spertino/%28sezione%29/curriculum>

Filippo Spertino -> Ingegnere Elettrico

(parole cercate:

CARATTERISTICHE CORRENTE-TENSIONE CELLE SOLARI MODULO FOTOVOLTAICO

Progettazione, sviluppo, collaudo impianti fotovoltaici autonomi connessi rete

SISTEMI IBRIDI PER ALIMENTAZIONE UTENTI RESIDENZIALI: DISPACCIAMENTO MULTIOBIETTIVO RETE/FOTOVOLTAICO

"sistemi automatici acquisizione dati"

Taratura strumenti misura.

Sistemi Elettrici Energia

Sistemi Elettrici Industriali

eolico

impianti fotovoltaici

Conversione fotovoltaica energia solare

turbina eolica

)

INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Electronic micro and nanosystems -> 50.66283 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Sistemi elettronici -> 17.82494 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Devices and Technologies for Integrated Electronics and Optoelectronics -> 17.30682 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Progettazione analogica e di potenza -> 7.48467 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Automazione -> 4.40276 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Embedded systems -> 1.20464 %

Progetto Intelligenza artificiale sviluppato da Teodoro Montanaro (matricola 188924)

INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Fiber optic systems -> 0.58464 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Microelectronics -> 0.19656 %
 FISICA DEI SISTEMI COMPLESSI (PHYSICS OF COMPLEX SYSTEMS) Magistrale: Unico -> 0.18023 %
 INGEGNERIA DEL CINEMA E DEI MEZZI DI COMUNICAZIONE Unico triennale -> 0.13041 %
 INGEGNERIA TELEMATICA (COMPUTER AND COMMUNICATION NETWORKS ENGINEERING) Magistrale: Unico -> 0.00540 %
 INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Information theory and signal processing applications -> 0.00498 %
 INGEGNERIA DEL CINEMA E DEI MEZZI DI COMUNICAZIONE Magistrale: Eurecom, Web Engineering, Multimedia -> 0.00316 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Embedded systems -> 0.00279 %
 INGEGNERIA MECCATRONICA (MECHATRONIC ENGINEERING) Magistrale: Unico -> 0.00254 %
 INGEGNERIA ELETTRONICA Unico Triennale -> 0.00123 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Software e sistemi digitali -> 0.00047 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Communication Systems Engineering, Eurecom -> 0.00032 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Wireless Systems Design -> 0.00030 %
 INGEGNERIA FISICA Unico triennale -> 0.00011 %
 INGEGNERIA DELLE TELECOMUNICAZIONI Unico Triennale -> 0.00009 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Progettazione a RF -> 0.00005 %
 INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Wireless systems -> 0.00003 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Reti -> 0.00001 %
 NANOTECNOLOGIE PER LE ICT (NANOTECHNOLOGIES FOR ICTs) Magistrale: Unico -> 0.00001 %
 INGEGNERIA INFORMATICA Unico triennale -> 0.00001 %
 INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Multimedia e reti -> 0.00000 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Software -> 0.00000 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Grafica e computer animation -> 0.00000 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Multimedia -> 0.00000 %

3. www.telematica.polito.it/oldsite/Laurea_Telematica/storia_valentina.html

Valentina -> Ingegnere Telematico

(parole cercate:

Cercavo una facoltà che mi desse possibilità di applicare matematica e fisica a qualcosa di reale, pratico, concreto. Cresciuta in mezzo a racconti di satelliti, ponti radio, reti cellulari e camere anecoiche.

comunicare! Silicon Valley e Cisco. high tech. cutting edge technologies.

)

INGEGNERIA MECCATRONICA (MECHATRONIC ENGINEERING) Magistrale: Unico -> 13.98220 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Embedded systems -> 11.36646 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Embedded systems -> 11.11263 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Electronic micro and nanosystems -> 9.89080 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Software -> 9.87359 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Sistemi elettronici -> 8.51839 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Software e sistemi digitali -> 7.92468 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Microelectronics -> 7.91608 %
 INGEGNERIA DEL CINEMA E DEI MEZZI DI COMUNICAZIONE Magistrale: Eurecom, Web Engineering, Multimedia -> 6.39309 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Grafica e computer animation -> 6.07043 %
 INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Multimedia e reti -> 5.77788 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Devices and Technologies for Integrated Electronics and Optoelectronics -> 0.21488 %
 FISICA DEI SISTEMI COMPLESSI (PHYSICS OF COMPLEX SYSTEMS) Magistrale: Unico -> 0.20360 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Communication Systems Engineering, Eurecom -> 0.16169 %
 INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Information theory and signal processing applications -> 0.16017 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Progettazione analogica e di potenza -> 0.15358 %
 INGEGNERIA FISICA Unico triennale -> 0.14394 %

Progetto Intelligenza artificiale sviluppato da Teodoro Montanaro (matricola 188924)

INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Fiber optic systems -> 0.13171 %
INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Reti -> 0.00175 %
INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Multimedia -> 0.00153 %
INGEGNERIA TELEMATICA (COMPUTER AND COMMUNICATION NETWORKS ENGINEERING) Magistrale: Unico -> 0.00069 %
INGEGNERIA INFORMATICA Unico triennale -> 0.00009 %
INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Wireless Systems Design -> 0.00004 %
INGEGNERIA DEL CINEMA E DEI MEZZI DI COMUNICAZIONE Unico triennale -> 0.00003 %
INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Progettazione a RF -> 0.00003 %
INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Wireless systems -> 0.00002 %
INGEGNERIA ELETTRONICA Unico Triennale -> 0.00002 %
NANOTECNOLOGIE PER LE ICT (NANOTECHNOLOGIES FOR ICTs) Magistrale: Unico -> 0.00001 %
INGEGNERIA DELLE TELECOMUNICAZIONI Unico Triennale -> 0.00000 %
INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Automazione -> 0.00000 %

4. <http://polimage.polito.it/~maurizio/ita/cv.html>

Maurizio Tranchero -> Ingegnere Elettronico

(parole cercate:

Circuiti stampati

Codesign

Tool sviluppo.

Sistemi asincroni

desincronizzazione processori.

Progettazione digitale

dispositivi silicio.

Progettazione analogica;

Sensori, trasduttori e circuiti di condizionamento analogici;

CAD per circuiti stampati

Coprogettazione e cosimulazione di sistemi misti HW/SW, Codesimulink;

back-end per i circuiti digitali,

Sintesi circuiti digitali

programmazione oggetti in C++

Matlab(TM)

Simulink;

UML

Linguaggi scripting

UNIX

Sistemi embedded

firmware micro-controllori.

)

Progetto Intelligenza artificiale sviluppato da Teodoro Montanaro (matricola 188924)

INGEGNERIA DEL CINEMA E DEI MEZZI DI COMUNICAZIONE Unico triennale -> 66.76577 %
 INGEGNERIA TELEMATICA (COMPUTER AND COMMUNICATION NETWORKS ENGINEERING) Magistrale: Unico -> 33.22116 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Automazione -> 0.00772 %
 INGEGNERIA DEL CINEMA E DEI MEZZI DI COMUNICAZIONE Magistrale: Eurecom, Web Engineering, Multimedia -> 0.00506 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Progettazione analogica e di potenza -> 0.00020 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Communication Systems Engineering, Eurecom -> 0.00003 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Devices and Technologies for Integrated Electronics and Optoelectronics -> 0.00003 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Wireless Systems Design -> 0.00002 %
 INGEGNERIA MECCATRONICA (MECHATRONIC ENGINEERING) Magistrale: Unico -> 0.00001 %
 INGEGNERIA FISICA Unico triennale -> 0.00000 %
 INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Information theory and signal processing applications -> 0.00000 %
 INGEGNERIA DELLE TELECOMUNICAZIONI Unico Triennale -> 0.00000 %
 FISICA DEI SISTEMI COMPLESSI (PHYSICS OF COMPLEX SYSTEMS) Magistrale: Unico -> 0.00000 %
 INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Wireless systems -> 0.00000 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Electronic micro and nanosystems -> 0.00000 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Embedded systems -> 0.00000 %
 INGEGNERIA ELETTRONICA Unico Triennale -> 0.00000 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Progettazione a RF -> 0.00000 %
 INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Fiber optic systems -> 0.00000 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Reti -> 0.00000 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Sistemi elettronici -> 0.00000 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Software -> 0.00000 %
 INGEGNERIA INFORMATICA Unico triennale -> 0.00000 %
 INGEGNERIA ELETTRONICA (ELECTRONIC ENGINEERING) Magistrale: Microelectronics -> 0.00000 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Software e sistemi digitali -> 0.00000 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Embedded systems -> 0.00000 %
 INGEGNERIA DELLE TELECOMUNICAZIONI (TELECOMMUNICATIONS ENGINEERING) Magistrale: Multimedia e reti -> 0.00000 %
 NANOTECNOLOGIE PER LE ICT (NANOTECHNOLOGIES FOR ICTs) Magistrale: Unico -> 0.00000 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Grafica e computer animation -> 0.00000 %
 INGEGNERIA INFORMATICA (COMPUTER ENGINEERING) Magistrale: Multimedia -> 0.00000 %

8) Motivi per cui ha funzionato solo 2 volte su 4:

Sicuramente la quantità di dati presenti nel database non è sufficientemente grande da poter fare una buona stima!

Ed inoltre potrebbe essere necessario ricontrollare tutti i pesi mettendo bene a confronto il peso per ogni ordinamento (va bene che la parola Informatica ha un peso quasi identico in Ingegneria Informatica e ingegneria del Cinema, ma la parola andrebbe pesata diversamente nell'ottica del fatto che chi cerca informatica molto probabilmente è poco interessata a Ingegneria del Cinema)

9) Problemi irrisolti

- a) Tra i sinonimi non viene considerata la possibilità che una parola possa essere singolare o plurale

[Per tutte le eccezioni si veda il sito <http://www.itg-rondani.it/dida/italiano/il%20nome/6.htm>]

Ed inoltre non vengono considerate tutte le possibili varianti grammaticali applicabili alle singole parole.

- b) Non è stato cercato un algoritmo di autoapprendimento che possa identificare le parole più importanti per un corso di laurea!

In realtà, visto il modo in cui le descrizioni dei corsi di Laurea sono scritti basterebbe andare a prendere tutte le parole più ricorrenti all'interno del testo, escludendo ovviamente gli articoli, le preposizioni, ecc ... !

- c) Ho dovuto mettere un limite (max 50) sul numero di parole da cercare perché dovendo effettuare tante connessioni al server EuroWordNet ed essendo su un server web, il rischio è che il server ci butti fuori a causa del troppo tempo dell'elaborazione (generalmente dopo 60 secondi il server ci butta fuori!)
- d) Bisognerebbe fare un po' più attenzioni ai problemi legati alla semantica: se ho la parola dispositivo e a seguire la parola mobile, il peso da attribuire è diverso dalla somma dei pesi delle 2 parole!

In realtà questo è stato già fatto, ma bisognerebbe rianalizzare bene i dati inseriti nel database in modo da avere una base dati un po' più solida (bisognerebbe guardare tutte le parole inserite e vedere se alcune di queste potrebbero avere qualche possibile legame da inserire.

10) Possibili sviluppi futuri

- a) Prendere tutti i dati presenti nel database ricontrollando tutte le relazioni con gli orientamenti e tutti i pesi via via assegnati.

In particolare sarebbe utile fare un tuning dei dati e dei pesi in modo da cercare un buon compromesso tra la reale relazione esistente tra parola e orientamento, e il peso che invece sarebbe meglio attribuire per i nostri scopi.

- b) Si potrebbe migliorare il sistema in modo che le relazioni e i pesi migliorino o vengano comunque modificati in base a dei feedback dell'utente che fa la ricerca (vedere tecniche di apprendimento per rinforzo)