

# Edit distance

Diego Biagini

July 9, 2019

# **1 Introduzione**

Un'operazione che viene frequentemente eseguita in programmi di vario tipo è ,data una parola sbagliata oppure incompleta, trovare la parola più vicina ad essa.

Esistono numerosi approcci per risolvere questo problema, tutti questi si basano sul confronto della parola voluta con un dizionario delle parole possibili.

Tra queste parole è possibile trovare quella più vicina attraverso la cosiddetta edit-distance, distanza di editing, più questa è piccola più le parole sono simili tra loro.

Eseguire questa operazione per ogni parola possibile non è molto raccomandabile dato il grande numero di parole in un dizionario.

Inoltre la ricerca della parola più vicina viene usata molto spesso in applicazioni real time, come suggerimenti di autocompletamento, è quindi necessario che abbia buoni tempi di esecuzione, anche minori di un secondo.

Per raggiungere questo obiettivo è necessario diminuire il numero di parole con cui eseguiremo edit-distance, garantendo però che la parola più vicina sia considerata tra questi confronti.

## **2 Cenni teorici**

### **2.1 Edit distance**

### **2.2 Intersezione di n-gram**

### **2.3 Coefficiente di Jaccard**

## **3 Esperimenti svolti**

## **4 Documentazione del codice**

## **5 Risultati sperimentali**

## **6 Analisi e conclusioni**