

UNIVERSITÀ DI PADOVA

SEMINARIO STUDENTI

Un Modello Geometrico per l'Information Retrieval

Autore:

Maria Maistro

Matricola:

1039607

4 febbraio 2014

Abstract:

In questo seminario presento il concetto di Information Retrieval e parlo delle sue applicazioni soprattutto per quanto riguarda l'ambito dei motori di ricerca. Introduco le definizioni di base: quella di documento, rilevanza ed esigenza informativa. Quindi descrivo brevemente la struttura di un sistema di Information Retrieval e le difficoltà principali che si incontrano nella costruzione di tali sistemi.

Lo scopo del seminario è di presentare un modello geometrico per l'Information Retrieval; quindi inizio illustrando la definizione di modello e di metafora, successivamente descrivo le sue principali componenti: descrittori, interrogazione e funzione di reperimento. Inoltre spiego sinteticamente in cosa consistono i modelli logico, probabilistico e linguistico che insieme a quello geometrico sono i principali modelli di Information Retrieval.

Infine mi concentro sul modello vettoriale e descrivo dettagliatamente la sua metafora: descrittori, interrogazione e documenti sono rappresentati da vettori o combinazioni di vettori mentre la funzione di reperimento è il prodotto scalare. In particolare mi soffermo sui vari schemi di pesatura per i descrittori nei documenti (binario, TF, IDF, TFIDF) e sulle diverse normalizzazioni della funzione di reperimento (metodo del coseno, del peso massimo e del perno). In seguito espongo il concetto di correlazione tra descrittori e conseguentemente definisco la matrice di correlazione e il problema del mismatch. Infine concludo parlando brevemente dei costi computazionali di tale modello e di alcuni semplici accorgimenti per limitarli.

Riferimenti bibliografici

- [1] Melucci, Massimo (2013), *Information Retrieval: Metodi e Modelli per i Motori di Ricerca*, FrancoAngeli, Milano.