

Capitolo 1

Introduzione

Con questa tesi si ha come obbiettivo di studiare e analizzare le varie tecniche di spam detection e inoltre analizzare le tecniche online. Oltre a questo le tecniche vengono classificate sulla base dei segnali che utilizzano. Il fattore chiave è che non ci sono, o meglio sono poche, al momento tecniche online di spam detection, ovvero tecniche che rilevano lo spam durante la frase di crawling. Infatti quasi tutti i metodi tentano di fare il crawling dell'intero web e successivamente classificare le pagine in spam o buone. Introduzione alla tesi Perché viene fatta la tesi Dove viene fatta Problemi relativi Soluzioni

1.1 Web spam

Con il termine web spamming si fa riferimento a tutti i metodi che tentano di manipolare gli algoritmi di ranking dei motori di ricerca per aumentare il valore di ranking di alcune pagine rispetto ad altre[1]. Dato il numero esorbitante di pagine che vengono create e pubblicate sul web, gli utenti competono per poter far comparire le proprie pagine tra le prime dei risultati di una query. Il fenomeno dello spamming o spamindexing comporta che la qualità delle ricerche decrementa, gli utenti tenderanno ad utilizzare altri motori di ricerca, l'indicizzazione di pagine che non sono utili e l'au-

mento del costo delle operazioni di query ed infine la causa di malware e reindirizzamento verso contenuto per adulti[2].

Bibliografia

- [1] Zoltan Gyongyi and Hector Garcia-Molina. Web spam taxonomy. Technical Report 2004-25, Stanford InfoLab, March 2004.
- [2] Nikita Spirin and Jiawei Han. Survey on web spam detection: Principles and algorithms. *SIGKDD Explor. Newsl.*, 13(2):50–64, May 2012.