



Allegato 1: Specifiche del progetto di ricerca di cui all'art. 4

Titolo: *“Ricerca, sperimentazione e sviluppo di algoritmi per applicazioni di mining web finalizzati all’antifrode ed al Marketing.”*

SSD: **ING-Inf/03**

Tutor: **Prof. Ciro D’Elia**

Abstract: Il contesto della ricerca è quello del mining in ambito WEB, tramite algoritmi studiati ad hoc e con l’utilizzo delle API’s fornite dai social network che verranno scelti come target del lavoro. Gli algoritmi dovranno essere in grado di estrarre dal WEB le informazioni basilari necessarie a Classificare contenuti WEB e ad individuare frodi e/o fenomeni di interesse per il marketing delle aziende, secondo una ontologia fornita che verrà raffinata nel corso delle ricerche.

Obiettivi: Obiettivo della ricerca è quello di realizzare algoritmi di estrazione di contenuti informativi da risorse web per applicazioni antifrode. Il web è potenzialmente una fonte inesauribile di informazioni, pertanto spesso le frodi informatiche possono essere prevenute e/o individuate esaminando l’attività pubblica degli utenti in rete. Tali informazioni devono tuttavia essere opportunamente elaborate ed interpretate: essendo esse stesse per propria natura destrutturate, ed essendo scritte essenzialmente in linguaggio naturale e pensate per essere fruite da persone, non si prestano ad essere processate in maniera immediata da sistemi automatici. Si pensi a scopo esemplificativo ai forum, nei quali, ad esclusione dell’appartenenza di un contenuto ad una specifica sezione, non si ha alcuna struttura nei dati, quindi non è immediata la creazione di un sistema per usufruire delle informazioni scritte nel forum stesso senza l’utilizzo di appositi algoritmi di data mining. L’aspetto contestuale di organizzazione dell’informazione che caratterizza il concetto di forum è stato poi negli ultimi anni ulteriormente esteso grazie ai social networks, all’interno dei quali non solo sono disponibili informazioni strutturate legate ai contenuti ma è anche possibile individuare un contesto più articolato grazie a dati quali ad esempio l’autore dei contenuti stessi, la sua rete, i tag, gli oggetti multimediali pubblicati, i suoi interessi ed altro ancora.



UNIVERSITÀ DEGLI STUDI
DI CASSINO E DEL LAZIO MERIDIONALE

Obiettivo è quindi quello di studiare e realizzare un algoritmo di data mining pensato appositamente per trattare queste tipologie di dato destrutturato. Un possibile approccio è quello di sfruttare il concetto di tassonomia, organizzando quindi il dato in classi gerarchiche via via sempre più specializzate in modo da caratterizzare l'informazione in maniera efficace, concentrandosi nello specifico nell'ambito delle frodi informatiche e degli scenari tipici ad esse legati. La tassonomia rappresenta uno strumento efficace di rappresentazione di informazioni, soprattutto nel caso in esame dove i contenuti con cui si ha a che fare sono estremamente eterogenei, e per i quali quindi si rende necessario un processo di classificazione. In particolare lo scopo è quello di utilizzare una tassonomia specifica che sfrutti da un lato i concetti basilari che caratterizzano tipicamente una frode informatica, e poi cali questo aspetto nel contesto dell'attività degli utenti in rete. Inoltre la ricerca sarà volta ad Algoritmi che, devono essere in grado di raffinare il processo di ricerca e arricchire la tassonomia a partire dai dagli esiti delle ricerche precedenti, e dalla loro valutazione da parte dell'utente.

Risultati
attesi: Gli algoritmi da realizzare dovranno quindi, a partire da una chiave di ricerca (nome / cognome / email ecc.), restituire informazioni strutturate a valore aggiunto per l'utente analista, come ad esempio un indice di affidabilità, la sua rete di contatti, allarmi ecc