

Text Mining: uno strumento al servizio degli operatori di Sicurezza ed Intelligence

Alessandro Zanasi
TEMIS Text Mining Solutions VP
Universita' di Modena e Reggio Emilia
Alessandro.zanasi@temis-group.com
Tel.: +349 41 31 718

Con il recente tragico salto di qualità del terrorismo internazionale torna di grande attualità il tema dell'impiego delle tecnologie informatiche per l'identificazione dei rischi e per la difesa nazionale. Un impiego che vede in prima linea le tecnologie d'intelligence e di text mining, le stesse che stanno oggi progressivamente entrando nell'uso comune delle aziende, per l'analisi dei comportamenti dei clienti o l'individuazione delle nuove opportunità di business.

Nell'intelligence effettuata per scopi militari, l'IT (Information Technology) interviene essenzialmente su tre livelli.

- Il primo riguarda la capacità di raccolta dei dati; intercettatori su reti telematiche, sensori satellitari, aerei spia possono monitorare virtualmente ogni cosa, raccogliendo grandi volumi di dati.
- Il secondo riguarda la capacità di analizzare queste moli di dati in maniera automatica. Sistemi avanzati di comando, controllo, comunicazione ed elaborazione (C4) danno un senso ai dati raccolti. È qui che si sono avuti i maggiori sviluppi, grazie ai sistemi di data e text mining che permettono di superare il problema dell'eccessiva informazione (*information overload*), estraendo la conoscenza che è insita in questi dati senza obbligarne l'analista a dover avanzare ipotesi sul loro contenuto.
- Il terzo consiste infine nella capacità di agire in particolare usando conoscenza IT per entrare in sistemi protetti.

Vediamo nel seguito uno degli strumenti di IT citati precedentemente e le sue applicazioni a casi di intelligence aziendale. Le applicazioni militari saranno facilmente intuibili.

1. Il text mining

Il text mining è lo strumento che permette di trattare i documenti con strumenti di analisi automatica. Questi strumenti variano considerevolmente tra loro ma, in generale, riassumono e categorizzano i documenti, identificano la lingua in cui sono scritti, estraggono concetti chiave, nomi propri e frasi con più parole, ne riportano frequenze, classificano un documento in funzione della rilevanza rispetto a uno specifico argomento.

Alcuni incorporano capacità di Web-crawling (ovvero di recupero automatico di documenti sulla rete internet), estraggono dati in vari formati, collegano le informazioni tra di loro in relazioni spaziali o temporali, scoprono legami o catene di informazioni legate fra di loro, raggruppano documenti in funzione del loro contenuto, effettuano analisi incrociate e permettono l'utilizzo congiunto di package statistici.

Attraverso il text mining si possono analizzare volumi immensi d'informazione, sia in tempo reale sia in differita e si possono identificare relazioni e strutture che altrimenti sfuggirebbero alla capacità analitica dell'essere umano.

Cosa è possibile scoprire con il text mining applicato a problemi di sicurezza nazionale?

Le elites politiche esprimono idee e pensieri sotto l'influenza di analisi di altissimo livello, di briefing militari e diplomatici, di gruppi di lavoro interagenti incaricati della formulazione della politica nazionale, di rapporti di commissioni e di altre fonti ricche d'informazione.

Queste idee e pensieri da loro espressi contengono perciò tracce del processo articolato attraverso il quale si va costruendo e palesando una decisione strategica.

Ma allora è vero anche il contrario: attraverso l'analisi contemporanea di molte fonti relative al medesimo argomento si possono rivelare strutture e legami che permettono di ricostruire le linee guida attraverso le quali viene pianificata una certa politica e di conseguenza prevedere le decisioni future.

Molte dichiarazioni di leader politici, per esempio, sono spesso prevedibili in anticipo attraverso osservazioni casuali fatte dai rappresentanti del loro staff.

A seconda dell'ambiente, le fonti d'informazioni più ricche possono essere all'interno di due o tre cerchi concentrici di persone che si trovano intorno al responsabile di un gruppo.

Questi cerchi concentrici che circondano i leaders sono composti da coloro che preparano i briefing decisionali, frequentano gli incontri dei gruppi di lavoro, preparano le bozze che indirizzano i decision maker e che strutturano la politica da seguire.

Bill Flynt, tenente colonnello dell'esercito USA, in un articolo apparso su Military Review del 7 Agosto 2000 esamina l'applicazione del processo di text mining al libro Unrestricted Warfare, scritto da due colonnelli dell'esercito cinese: Qio Ling e Wang Xiangsui. Sebbene i risultati migliori del text mining si abbiano quando le fonti informative sono in numero elevato, già in questo caso Flynt ne dimostra l'utilità per riconoscere, attraverso l'analisi automatica delle idee espresse dagli autori, le diverse minacce emergenti nelle nuove forme di guerra, i mezzi utilizzati per realizzarle, gli obiettivi e i fini strategici che si pensa di raggiungere. L'analisi mette inoltre in evidenza le relazioni che gli autori individuano come essenziali per raggiungere certi obiettivi strategici. Flynt sottolinea l'importanza del text mining per individuare le prospettive strategiche di un certo soggetto sia che si tratti dell'esercito cinese sia che si tratti, secondo le sue testuali parole, di un terrorista come Unabomber o dell'autore di una pagina Web.

2. L'informazione elaborata

L'informazione si trova contenuta nelle migliaia di banche dati accessibili on line in tutto il mondo, nonché sui siti Web il cui numero sta crescendo in maniera esponenziale. Quest'informazione consiste in pagine di tesi, di memorie, di pubblicazioni scientifiche relative a un numero imponente di argomenti facilmente accessibili, in testo libero, online. Alcune sono gratuite, anche se la maggioranza di quelle più interessanti sono a pagamento, raccolte attraverso i servizi di qualche infoprovider, come EINS (European Information Network Services), Factiva, Dialog o Lexis/Nexis.

I servizi di intelligence (militare e non) sono stati i primi a interessarsi a strumenti e metodologie in grado di elaborare il contenuto di queste data bank per estrarne non tanto informazioni puntuali (che si troverebbero senza grossi problemi) ma, assai più intelligentemente, informazioni di livello superiore, quelle che consistono nelle correlazioni delle informazioni di base tra di loro: sia che riguardino luoghi, dichiarazioni, nomi di persone o concetti.

Gli USA cominciarono, attraverso il progetto DARPA (da cui nacque poi Internet), a interessarsi di questi sistemi fin dagli anni '70. L'obiettivo a quell'epoca era di estrarre le informazioni sensibili da grandi moli di dati testuali in maniera automatica. Inizialmente l'obiettivo era solo di trattare i messaggi (brevi) che arrivavano dall'esterno e di farli arrivare alle persone che più ne erano interessate (attraverso tecniche di filtering e routing). Ma poi l'interesse si spostò verso sistemi che potevano andare da soli a cercare le informazioni su database esterni.

Con l'esplosione del fenomeno di Internet nella seconda metà degli anni '90, si è aperto un nuovo fronte per gli operatori di Intelligence.

Infatti non tutti gli internauti sono coscienti del fatto che, anche se due amici che si scambiano e-mail abitano nello stesso palazzo, è sufficiente che uno di loro utilizzi come server di posta elettronica quello di un'azienda localizzata in un'altra nazione (Compuserve, Freesurf, Yahoo...) perché il loro carteggio sia disponibile all'analisi linguistica avanzata dei tool di text mining dei responsabili della sicurezza di quel Paese.

Con batterie di filtri informatici e di analizzatori semantici, diviene possibile creare in maniera automatica, da una raccolta di documenti immensa ed eterogenea, costituita da e-mail, siti Web e banche dati, nuove informazioni come, per esempio, le tendenze d'opinione di un certo gruppo di individui.

Per ora si conoscono bene le applicazioni che di queste tecnologie stanno facendo aziende come IBM, Telecom Italia Mobile, Unilever, Aerospaziale, Michelin o Pfizer. Fondamentalmente queste aziende, grazie a questi sistemi, conoscono esattamente i trend evolutivi di qualunque tecnologia o strategia dei loro competitors senza essere obbligati a fare un noioso lavoro di selezione manuale e di lettura o, peggio, illegale.

Alla fine degli anni '90 sono nati prodotti come Text Knowledge Miner, di IBM, Semio Map, di Semio, DR-LINK, di Insight Discoverer, di Temis. Prodotti di nuova generazione, che cercano di implementare l'analisi semantica, ovvero la capacità di 'capire' il significato dei documenti, risolvendo i problemi nascenti dalle ambiguità del linguaggio, delle sue forme retoriche, delle antonomasie, metafore, anafore e paratassi.

Possiamo dire che l'informatica, dopo aver automatizzato negli anni '70 ed '80 il settore secondario, ovvero le attività dedicate alla produzione, negli anni '90 il terziario, ovvero i servizi, come banche e assicurazioni, si è lanciata con forza con l'avvio del nuovo millennio nel terziario avanzato, o quaternario: l'automazione, attraverso il text mining, di tutto ciò che riguarda il marketing, la consulenza, l'entertainment.

Nel 2000 un gruppo di specialisti italiani, francesi, tedeschi ed americani hanno lasciato l'azienda per cui lavoravano (IBM) per fondare TEMIS, una società specializzata nello sviluppo di software di text mining e realizzando un sistema di intelligence usato nel mondo militare come nel civile. Nel civile esso è utilizzato per individuare opportunità di business per piccole e medie imprese. La tecnologia di text mining semantico online di Temis consente di identificare queste opportunità nel mare magnum di milioni, miliardi di pagine disponibili su siti Web e all'interno di migliaia di banche dati online.

Faremo riferimento nel seguito agli approcci innovativi realizzati da questa società'.

3. Text Mining: come funziona?

Introduzione

Text mining è una forma particolare di data mining dove i dati, consistenti in testi liberi, sono destrutturati. A causa della presenza del linguaggio naturale, lo step di preparazione dati è più lungo del solito e richiede una fase di preprocessing linguistico [BRZ99],[GG98] per sciogliere, almeno parzialmente, le ambiguità legate alla comprensione del significato. In questo senso il text mining estende il concetto di *content analysis* [We90], che non prevedeva la possibilità di trattare il testo prima di analizzarlo, nella direzione mostrata da [Man97].

L'obiettivo rimane, comunque, lo stesso: scoprire la conoscenza nascosta nei dati, in questo caso documenti, senza definire in anticipo l'argomento della ricerca o definendolo in maniera generica (cfr.1.2.4 Fase di scoperta di regole).

La soluzione proposta, pur avvantaggiandosi dei tools più avanzati disponibili ed integrando, ad esempio, diversi motori di ricerca, differisce da altre soluzioni dirette ad offrire modi più efficienti di effettuare ricerche nei databases testuali, una volta che sia stato definito l'argomento oggetto della ricerca. Queste ultime soluzioni si basano infatti su di un diverso approccio, diretto ad offrire migliori capacità di ricerca e navigazione, basandosi, ad esempio, sulla considerazione che molti *corpora*, come le directories di Internet, le librerie digitali e i databases di brevetti, sono organizzati in gerarchie (tassonomie). Il concetto chiave in questo trend di ricerca sta nel definire tecniche che permettano di separare, in ogni nodo della tassonomia, le parole discriminanti da quelle che rappresentano *rumore* (cfr.[CDAR98]).

Fase di preprocessing linguistico

Utilizzando dizionari elettronici, taggers sintattici, motori di lemmatizzazione, la soluzione proposta:

- risolve le principali ambiguità legate alla lingua; in questa fase, ad esempio, le parole:
 - *record*, nella frase inglese ‘we record the record’, - *couvent*, nella frase francese ‘les poules du couvent couvent’, - *pesca*, nella frase italiana ‘pesca la pesca’, sono riconosciute nel loro significato.
- lemmatizza parole/espressioni (“International Business Machines” è trasformato in “IBM”),
- indicizza automaticamente i documenti, associando loro i concetti chiave ivi contenuti.

Fase di data mining

Ove i dati sono trattati seguendo l’approccio classico del data mining (cfr.[Za97]), individuando connessioni/legami/somiglianze tra i diversi temi individuati.

Fase di scoperta di regole

Possono essere evidenziati, attraverso grafici e statistiche, una serie di temi di particolare interesse ed i loro collegamenti.

I temi ed i collegamenti possono riguardare:

- Nuove aree di business o di ricerca.
- Relazioni tra attività aziendali e/o aree di ricerca.
- Trends temporali.
- Concorrenti e loro attività.
- Alleanze/Partnerships/M&A’s.
- Previsioni dichiarate sul mercato.
- Dichiarazioni di scienziati/esperti/top executives.
- ...

anche se le parole precedenti (ad esempio *partnership* o il nome dell’azienda) non appaiono in maniera esplicita nei testi

Gli argomenti possono essere mutati a seconda delle esigenze (quindi possono essere reperite personalità politiche, managers, proprietari, azionisti, previsioni etc.etc.).

4. Un’Applicazione pratica di Text Mining: Competitive Intelligence

Overview

L’applicazione presentata, (*TEMIS Online Miner*), è una soluzione consistente in SW e servizi di consulenza, in grado di estrarre informazioni da grandi raccolte di documenti assicurando funzioni di ricerca, analisi, categorizzazione e navigazione. Tale soluzione mette in evidenza *segnali deboli*, invia messaggi e documenti alle persone in base al loro profilo (*alert system*), utilizza alcuni dei metodi più recenti di analisi automatica di dataset scaricati via rete [HQD91], [Hu96].

Un “precursore” di quello che prese il nome di Online Miner, è già usato in IBM [Za00] nella funzione di Market Intelligence per analizzare documenti provenienti da più di 3000 fonti stampa e da più di 70 siti web.

Temis Online Miner ha lo scopo di individuare opportunità/rischi presenti nel mercato e di permettere al management di reagirvi rapidamente. E’ stato poi implementato in un grande gruppo bancario italiano (SSB), in società operanti nel mondo delle telecomunicazioni (Telcal, TIM) e poi all’estero in varie altre realtà (Novartis, Dresdner Bank, Conoco, Renault...).

Il processo

Dopo aver approntato il server (inhouse o outsourced), si definiscono le tematiche di interesse e conseguentemente le fonti da cui selezionare e recuperare i documenti utili. In questa fase si definiscono, ad esempio, i profili aziendali di interesse (comprendenti le aziende controllate, i top executives, i maggiori azionisti), i temi aziendali o tecnologici, i segnali da individuare. In questa fase si definiscono pure i profili utenti in funzione della documentazione che dovranno ricevere.

Caricamento/Alimentazione dati

L'alimentazione dati è assicurata in maniera automatica e periodica. Le fonti dati sono, attualmente, di due tipi: agenzie stampa e siti web.

Le prime (fornite, ad esempio, da Reuters, Dow Jones e Bloomberg) sono spedite direttamente dall'info provider. I secondi sono ricercati e caricati attraverso appositi web crawlers.

Filtraggio ed analisi linguistica di documenti strutturati e non strutturati

Se i documenti da analizzare sono altamente strutturati (come nel caso dei brevetti) l'Online Miner, attraverso i suoi filtri, scompone tali documenti nelle loro componenti principali (Titolo, Abstract, Documento principale, Capitoli, Conclusioni, Bibliografia) e li riformatta in un formato omogeneo XML, riconoscendo e separando il contenuto strutturato da quello non strutturato.

Se i documenti non sono strutturati, attraverso l'analisi linguistica si riconoscono e classificano i temi presenti nei testi [DGS99].

Le tre componenti di base dell'analisi linguistica sono:

1. modello linguistico; fornisce la conoscenza grammaticale per scomporre le frasi nelle loro componenti di base: nomi, verbi, aggettivi, date etc., selezionando le nuove keywords dal testo presente;
2. dizionario generico di parole ed espressioni multi-parola, integrato con dizionari specifici per analizzare documenti tecnici;
3. motore di "estrazione di relazioni", per individuare relazioni come <soggetto X><verbo o espressione di significato><soggetto Y>, per individuare i cosiddetti "segnali"[GG98]. L'informazione estratta sarà automaticamente classificata in categorie quali:
 - Nomi di persone, organizzazioni e zone geografiche (come Dott.Ciampi, Repubblica Italiana, Roma, Italia);
 - Espressioni multiparola (Asset Liability Management, British Telecom);
 - Abbreviazioni (ALM, per Asset Liability Management; BT, per British Telecom);
 - Relazioni (Bill Gates-Presidente-Microsoft, Compaq-possiede-DEC);
 - Altro: date, valute, forme testuali di numeri, etc.

Una forma canonica sarà sempre assegnata ad ogni parola riconosciuta nel documento. Sarà il termine più esplicito, meno ambiguo tra le varianti che si riferiscono a questo soggetto. Quest'approccio permette di ridurre le varianti di un termine (in un documento *IBM* sarà associato ad *International_Business_Machines*, in un altro a *Industrie_Bolognesi_Meccaniche*).

Contenuto

Al termine dello step di analisi linguistica è creata una matrice che incrocia i documenti con le forme canoniche delle parole presenti nei documenti stessi. Questa matrice viene ora analizzata automaticamente creando, ad esempio, clusters di documenti simili. Ogni documento sarà così assegnato ad uno ed un solo cluster (quello che contiene i documenti a lui più simili).

Ogni cluster avrà un 'titolo', costruito attraverso le parole più importanti presenti al suo interno, che indicherà il contenuto tematico del cluster.

Attraverso l'utilizzo di barre colorate sono mostrate le relazioni tra clusters. Queste relazioni indicano che due aree, apparentemente non aventi elementi in comune, sono in verità legate tra di loro in quanto i due settori (aziendali o tecnologici) condividono "qualcosa": ad esempio, può essere una tecnologia nuova, risultante dalla sinergia tra due tecnologie diverse. Può anche essere mostrata una relazione tra un'area geografica e una tecnologia, od una capacità aziendale, di particolare valore.

Scoperta di regole

L'Online Miner riconosce i segnali di cambiamento nell'ambiente competitivo come : dichiarazioni di top executives, partnerships, alleanze, previsioni pubblicate sui media.

Individua tali segnali dopo averli trasformati in regole.

Esempio:

SE

leggi <azienda_X><frase_Segnale><azienda_Y>

ALLORA

scrivi "Segnale (X,Y)".

Customer Intelligence

L'approccio precedentemente visto in casi di Competitive Intelligence puo' essere specularmente applicato a casi di CRM. In tali casi i dati saranno di provenienza emails, chatlines, forums, newsgroups etc.etc.. Le regole estratte riguardano i sentimenti di soddisfazione o meno dei clienti, oltre a quelle gia' presentate nei casi di Competitive Intelligence.

L'esempio presentato riguarda email inviate ad un Forum che raccoglie opinioni da parte di acquirenti di auto sull'auto da loro acquistata. L'applicazione permette di distinguere automaticamente email riguardanti richieste, ordini o lamentele. E queste ultime sono a loro volta separate automaticamente in gruppi a seconda delle ragioni della lamentela.

5. Effettiva necessità di analisi linguistica

E' tema dibattuto quello relativo all'effettiva utilità/necessità di un approccio *text mining* (utilizzando l'analisi linguistica/semantica) per l'analisi del contenuto delle banche dati destrutturate.

Attualmente sono effettivamente sul mercato molti prodotti che operano analisi dei contenuti basandosi solo su tecniche statistiche. Tale approccio ha alcuni vantaggi. Esso permette, ad esempio, di essere

- language independent (la statistica è indipendente dalla lingua e gli stessi tools possono essere applicati a lingue diverse),
- di maggior semplicità (non essendo necessarie competenze linguistiche, ad esempio di implementazione di grammatiche formali).

D'altra parte proprio questa semplicità comporta alcuni svantaggi (eliminabili solo a prezzo di una complessità del sistema spesso talmente elevata da annullare il vantaggio derivante dall'assenza di analisi linguistica):

- incapacità a sciogliere ambiguità (*pesca la pesca*)
- incapacità ad individuare sinonimi (*vettura-automobile*)
- incapacità a gestire sinonimi variabili a seconda del contesto ("matrimonio" tra due imprese inteso come "alleanza" e non come sacramento)
- necessità di grande sforzo per personalizzare il sistema il sistema (*training*)
- mancanza di trasparenza relativa alle tecniche statistiche utilizzate ed al "training" necessario al sistema [HL00].

In breve, la supposta maggior semplicità non pare essere sufficiente a giustificare l'assenza di un approccio che tenga conto delle possibilità offerte dalla linguistica [YL99]. E' indicativo che i più importanti servizi di intelligence militari (sia Americani che Europei) stiano da alcuni anni lavorando su quest'ultimo approccio, che coniuga la matematica con la linguistica.

AUTORE

Alessandro Zanasi (alessandro.zanasi@temis-group.com, a_zanasi@yahoo.it) è laureato in Ingegneria Nucleare all'Università di Bologna, specializzato in matematica applicata a problemi di intelligence all'Università di Paris VI e all'Università di Modena, ove tuttora insegna.

Iniziò la sua attività come Ufficiale dei Carabinieri, al Centro Carabinieri Investigazioni Scientifiche (CCIS) di Roma. Dopo un periodo come consulente tecnico per diverse organizzazioni statali, entrò in IBM dove, dopo alcuni incarichi in USA (San Jose, CA) e Francia (Paris), e' stato responsabile della funzione di Market Intelligence nell'area Sud Europa.

Nel 2000 fonda Temis, di cui e' VP della holding e Direttore della filiale italiana.
Dal 2001 è Program Director (Int'l) di META Group, Inc., nell'area business intelligence.

Ha coordinato le attività del Bologna KDD Center, basato presso Cineca (Consorzio Interuniversitario per il calcolo automatico), ha tenuto corsi e conferenze all'Università Bocconi di Milano; all'Università Federale di Rio de Janeiro (UFRJ); alle Scuole di Guerra ed Accademie Militari Italiana, Francese e Brasiliana, alle Scuole dell'Arma dei Carabinieri; a workshop internazionali. Coordina progetti relativi a temi di Informazione Online ed Intelligence sia in Europa che in Nord e Sud America, ed e' attualmente consulente della Commissione Europea su tematiche di business intelligence.

Membro dell'IAFE (International Association of Financial Engineers), dell'European Advisory Board di SCIP (Society of Competitive Intelligence Professionals) e coordinatore del chapter SCIP di Bologna, dell'Associazione Internazionale di Polizia e dell'Associazione Nazionale dei Carabinieri, dell'Associazione Italiana per l'Intelligenza Artificiale, del comitato scientifico della rivista Scienza e Business. E' Socio d'onore dell'Associazione ItaSForum

E' autore/coautore di diverse pubblicazioni su argomenti di intelligence elettronica.

Bibliografia

Si rimanda il lettore interessato ad approfondimenti al libro

A.Zanasi - "Text Mining and its Applications to Intelligence, CRM and Knowledge Management" – 2005 – WIT Press

Articoli riguardanti il tema, citati o a cui si fa riferimento nel precedente articolo:

[BKC95] - Bologna KDD Center - <http://open.cineca.it/datamining>

[BRZ99] – R.Baeza-Yates, B.Ribeiro-Neto, N.Ziviani, 1999 – Text Operations, in "R.Baeza-Yates, B.Ribeiro-Neto: Modern Information Retrieval" – pp.163-190 – Addison Wesley

[BST00] – A.Berson, S.Smith, K.Thearling, 2000 – Building Data Mining Applications for CRM – pp.457-484 – McGrawHill

[CDAR98] – S.Chakrabarty, B.Dom, R.Agrawal, P.Raghavan, 1998 – Scalable feature selection, classification and signature generation for organizing large text databases into hierarchical topic taxonomies – The VLDB Journal (1998) 7:163-178

[DGS99] - Doerre, Gerstl, Seiffert, 1999 - Text Mining: Finding Nuggets in Mountains of Textual Data - KDD99 Proceedings - ACM

[GG98] - C.Gouffas, J.M.Granier, 1998 - The Death of Lady D - A Semiotic and Psycho-sociological Understanding of the Funeral Eulogy - ESOMAR Seminar on the Internet and Market Research

[GG96] - C.Gouffas, J.M.Granier, 1996 - Les mots de l'Entreprise: Analyse Textuelle Automatique et Semiotique - JADT Rome 96 Seminar

[HL00] – S.Hayward, A.Linden, 2000 – Gartner Group RAS Services – 6/6/2000

[HQD91] - C.Huot, L.Quoniam, H.Dou, 1991 - A new Method for Analyzing Downloaded Data for Strategic Decision - Scientometrics, Vol.22-No.2

[Hu96] - C.Huot, 1996 - IBM Technology Watch - IBM Internal Report

[Ma91] - J.F.Marcotorchino, 1991 - L'Analyse Factorielle Relationnelle (partie I et II) - Etude du CEMAP IBM France - N°MAP-003

[Ma99] – R.Mattison, 1999 - Web Warehousing and Knowledge Management, pp.337-354 - McGrawHill

[Man97] – I.Mani et al., 1997 – Towards Content-Based Browsing of Broadcast News Video – in Intelligent Multimedia Information Retrieval - Ed.AAAI Press

[We90] – R.P.Weber, 1990 – Basic Content Analysis – SAGE Publications

[YL99] – E.Younker, A.Linden, 1999 – CIO Update:Data Mining Applications of the next Decade –7/7/1999, Gartner Group Article

[Za97] - A.Zanasi, 1997 - Discovering Data Mining - Prentice Hall

[Za98] - A.Zanasi, 1998 - Competitive Intelligence Thru Data Mining Public Sources - Competitive Intelligence Review - Vol.9(1) - John Wiley & Sons, Inc.

[Za00] – A.Zanasi, 2000 – Web Mining through the Online Analyst - Data Mining 2000 Proceedings – Wessex Institute of Technology

[Za02] – A.Zanasi, 2002 (ed.) – Data Mining III – Wessex Institute of Technology

[Za02] – A.Zanasi, 2002 (ed.) – Applications of High Performance Computing – Wessex Institute of Technology