

Codice selezione: 12/2008

CURRICULUM VITAE *

DATI PERSONALI

Nome e Cognome: **Alessio Canzonetti**

Luogo e data di nascita: **Orbetello, 18-09-1972**

Residenza: _____

Tel.: _____

STUDI

Laurea in Economia c/o l'Università degli Studi di Roma "La Sapienza"

Data: 11 luglio 2001

Votazione: 104/110

Titolo della tesi: "I tratti caratteristici del lessico di frequenza del linguaggio economico-finanziario: differenze tra old e new economy"

in: Statistica

Relatore Prof. Sergio Bolasco

STUDI POST UNIVERSITARI

(dottorato di ricerca, corsi di specializzazione, di perfezionamento, borse di studio)

Tipo di corso: _____ **c/o:** _____

Durata: _____ **Anno:** _____

Titolo eventuale elaborato finale: _____

Titolo conseguito o periodo di frequenza: _____

Giudizio finale eventuale: _____

CORSI DI FORMAZIONE

Tipo di corso: Corso di formazione su software statistico – "Corso introduttivo all'utilizzo di R" c/o: Facoltà di Scienze Statistiche – Sapienza Università di Roma

Durata: 21 ore Anno: 2007

Titolo conseguito o periodo di frequenza: 26, 27 e 28 febbraio 2007

Tipo di corso: Corso di formazione su software statistico – "Statistica multivariata con R" c/o: Facoltà di Scienze Statistiche – Sapienza Università di Roma

Durata: 21 ore Anno: 2007

Titolo conseguito o periodo di frequenza: 25, 26 e 27 giugno 2007

PUBBLICAZIONI

1)

Autori: Canzonetti Alessio

Titolo: Information retrieval e analisi delle cooccorrenze per l'estrazione di informazione specifica da documentazione giuridica, in JADT 2008 - Actes des 9es Journées Internationales d'Analyse Statistique des Données Textuelles, pp.277-284, Presses Universitaires de Lyon, Lyon.

Curriculum vitae di Alessio Canzonetti

Data di pubblicazione: 2008

2)

Autori: Canzonetti Alessio

Titolo: Semantic classification and cooccurrences: a method for the rules production for the information extraction from textual data, in *CLADAG 2007 - Classification and Data Analysis 2007 - Book of Short Paper*, pp. 259-262, Macerata.

Data di pubblicazione: 2007

3)

Autori: Canzonetti Alessio

Titolo: Semantic classification and cooccurrences: a method for the rules production for the information extraction from textual data (versione estesa). In corso di pubblicazione in C. Lauro, F. Palumbo, M. Greenacre (eds.), *Data Analysis and Classification: from the exploratory to the confirmatory approach*, Springer series on "Studies in Classification, Data Analysis, and Knowledge Organization".

Data di pubblicazione: 2007

4)

Autori: Canzonetti A., Misuraca M., Romano M.C.

Titolo: Analyzing the language of everyday life: how Textual Statistics can support Time Use Surveys. The Italian Experience, in *Società Italiana di Statistica - Atti della XLIII Riunione Scientifica*, pp. 69-80, Torino.

Data di pubblicazione: 2006

5)

Autori: Canzonetti A., Capo F.M., Ruocco V.

Titolo: (2006), Estrazione di informazione da una base documentale dell'AGCM con il software Taltac2: un esempio di integrazione fra strumenti di text mining e tecniche di data mining, in *JADT 2006 - Actes des 8es Journées Internationales d'Analyse Statistique des Données Textuelles*, pp.245-252, Presses Universitaires de Franche-Comté, Besançon.

Data di pubblicazione: 2006

6)

Autori: Bolasco S., Canzonetti A., Capo F.M., della Ratta-Rinaldi F., Singh B.K.

Titolo: Understanding Text Mining: a pragmatic approach, in Sirmakessis S. (ed.), *Knowledge Mining*, pp. 31-50, Springer-Verlag, Berlin-Heidelberg.

Data di pubblicazione: 2005

7)

Autori: Bolasco S., Canzonetti A.

Titolo: Some insights into the evolution of 1990s' standard Italian using Text Mining techniques and automatic categorisation, in Vichi M., Monari P., Mignani S., Montanari A., (eds.), *New developments in classification and data analysis*, pp. 293-302, Springer-Verlag, Berlin-Heidelberg.

Curriculum vitae di Alessio Canzonetti

Data di pubblicazione: 2005

8)

Autori: Bolasco S., Canzonetti A., Capo F.M.

Titolo: Text Mining: uno strumento strategico per imprese e istituzioni, CISU, Roma. ✓

Data di pubblicazione: 2005

9)

Autori: Canzonetti Alessio

Titolo: La variabilità dei poliformi nel lessico di frequenza del linguaggio economico-finanziario, in *Annali del dipartimento di studi geoeconomici, linguistici, statistici, storici per l'analisi regionale 2003/2004*, Università degli Studi "La Sapienza" di Roma, pp. 337-350, Patron Editore, Bologna.

Data di pubblicazione: 2004

10)

Autori: Bolasco S., Baiocchi F., Canzonetti A., della Ratta-Rinaldi F., Feldman A.

Titolo: Applications, sectors and strategies of Text Mining, a first overall picture, in Sirmakessis S. (ed.) *Text mining and its applications*, pp. 37-52, Springer-Verlag, Heidelberg.

Data di pubblicazione: 2004

11)

Autori: Bolasco S., Canzonetti A.

Titolo: Quadro generale sugli istituti partecipanti al progetto, Valutazione dei corsi di aggiornamento, Le attività del progetto nelle istituti scolastici, in Morani R.M., Salustri M.C., Tais E. (a cura di), *Sapere i sapori: monitoraggio delle attività nelle scuole del Lazio*, pp. 23-108, Anicia, Roma.

Data di pubblicazione: 2004

12)

Autori: Bolasco S., Canzonetti A.

Titolo: Text mining as key to reading the Nineties' language in the daily "La Repubblica", *Atti del Convegno SIS 2003 – Analisi Statistica Multivariata per le Scienze Economico-Sociali, le Scienze Naturali e la Tecnologia*, Università di Napoli "Federico II", 9-11 Giugno 2003.

Data di pubblicazione: 2003

LAVORI (NOTE E RELAZIONI TECNICHE)

Autori: Bolasco S., Canzonetti A.

Titolo: Monitoraggio sull'attività didattica dei corsi di 2° semestre della formazione comune del Nuovo ordinamento (a.a. 2002/2003), in Università degli Studi di Roma "La Sapienza" – Nucleo di Valutazione d'Ateneo, *Relazione – Didattica e Ricerca: Attività 2003/2004*, Edizioni Kappa, Roma, pp. 147-160.

Data di pubblicazione: 2005

ESPERIENZE LAVORATIVE (partendo dalla più recente)

1)

Periodo: Aprile 2006 – Marzo 2008

Ditta o Ente: Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale della Facoltà di Economia di Sapienza - Università di Roma

Indirizzo: Via del Castro Laurenziano, 9 - 00161 Roma

Posizione: Assegnista di ricerca

Mansioni svolte e risultati conseguiti:

Il titolo della ricerca è "Statistica testuale: strumenti e tecniche integrate per processi di Data Text Mining; sviluppo di software per l'analisi dei dati testuali e di risorse statistico-linguistiche per il settore economico finanziario". È stata prestata particolare attenzione ai processi di estrazione di informazione dai testi e conseguente archiviazione in appositi database, da sfruttare quindi con tecniche di Data Mining. A questo scopo sono state messe a punto tecniche semi-automatiche di recupero dell'informazione desiderata che utilizzano grandezze desumibili direttamente dai testi. Queste tecniche comportano anche l'individuazione di regole, schemi di estrazione generalizzabili a più tipologie testuali (si vedano pubblicazioni 1, 2 e 3). Le conoscenze acquisite in fase di ricerca sono state implementate e testate attraverso il software TALTAC.

Inoltre, nel quadro di una convenzione tra l'ISTAT e la Facoltà di Economia il sottoscritto ha portato avanti un'analisi dei testi dei diari delle attività giornaliere di un campione della popolazione italiana raccolti nella ambito dell'indagine TUS (*Time Use Survey*). I diari raccolti sono relativi ad oltre 50.000 individui, per un numero di record pari a 1,5 milioni. Ogni record infatti contiene una singola attività svolta durante la giornata. Il sottoscritto ha pertanto dovuto approntare una ristrutturazione del dato al fine di ottenere un singolo record per ogni individuo, record che contenesse tutte le attività dell'intera giornata. Per ogni giornata (persona), erano disponibili anche informazioni strutturate. L'analisi ha poi coinvolto tecniche di estrazione di informazione per integrare le variabili già disponibili con variabili desunte dai testi. La matrice dati così arricchita è stata sottoposta ad una analisi fattoriale multipla su tabelle di contingenza con lo scopo di sintetizzare e porre in evidenza le attività giornaliere principali degli individui in età lavorativa, distinti per sesso, età e grado di istruzione (si veda pubblicazione 4).

Nell' a.a. 2007/2008 ha inoltre tenuto 40 ore di lezione nell'ambito del corso "Laboratorio avanzato di analisi dei dati e data mining" (docente: Prof. Tancredi) presso la Facoltà di Economia di Sapienza – Università di Roma. Ha inoltre svolgendo le lezioni di laboratorio informatico nell'ambito del corso "Metodi esplorativi e laboratorio di analisi dei dati" (docente: Prof.ssa Arezzo). Nei corsi vengono trattate tecniche fattoriali, *cluster analysis* e metodi di segmentazione, con l'ausilio di software per analisi statistica quali SPAD, SPSS e SAS (*Enterprise Miner* - modulo di Data Mining) a cui gli studenti vengono introdotti.

2)

Periodo: 2006 - 2008

Ditta o Ente: CITICORD di Sapienza – Università di Roma

Indirizzo: P.le Aldo Moro, 5 - Roma

Curriculum vitae di Alessio Canzonetti

Posizione: Docente

Mansioni svolte e risultati conseguiti:

Dal 2006 svolge attività di docenza nell'ambito dei corsi di formazione sul software per analisi statistica dei dati testuali e Text Mining TALTAC presso il CITICORD di Sapienza – Università di Roma. I corsi svolti nell'arco di questi due anni, sono stati 14, sia di livello introduttivo al software, sia di livello avanzato, ed hanno coinvolto oltre 100 partecipanti in totale provenienti dal oltre 40 dipartimenti in Italia.

3)

Periodo: 15 Luglio 2004 – 15 Settembre 2004

Ditta o Ente: Dipartimento di Scienze Umane dell'Università per Stranieri di Siena

Indirizzo: Piazza Carlo Rosselli 27/28- 53100 Siena

Posizione: Collaboratore di ricerca

Mansioni svolte e risultati conseguiti:

Nel 2004 ha svolto attività di collaborazione con il Dipartimento di Scienze Umane dell'Università per Stranieri di Siena avente per oggetto "Indagine per il monitoraggio della formazione mediante somministrazione ed elaborazione di questionari sottoposto agli studenti frequentanti il nuovo ordinamento didattico". Allo scopo è stata fornita consulenza per la costruzione dei tracciati record e per l'inserimento delle informazioni contenute nei questionari cartacei. Sul dato elettronico così ottenuto sono state operate analisi statistiche di tipo descrittivo. La collaborazione si è estesa anche ad uno studio delle caratteristiche dei contenuti di sette testi di didattica della lingua italiana per stranieri, i cui risultati verranno presentati in un volume di prossima pubblicazione. L'obiettivo era quello di giungere ad individuare le principali caratteristiche della tipologia testuale "materiale didattico", operando pertanto confronti con altre tipologie (linguaggio della stampa, politico, standard e parlato). Inoltre sono state analizzate alcune caratteristiche dei sette singoli testi (grandezze lessicometriche, linguaggio peculiare, semplicità del linguaggio adottato, caratteristiche grammaticali, coerenza nell'esposizione degli argomenti) allo scopo di individuare i diversi approcci alla didattica seguiti dalle sette opere.

4)

Periodo: 2004 – 2006

Ditta o Ente: BSD s.r.l.

Indirizzo: Via Trieste, 32 - 35121 Padova

Posizione: Collaboratore esterno

Mansioni svolte e risultati conseguiti:

Nel periodo 2004-2005 ha progettato e creato, per conto di BSD s.r.l. e Telecom Italia, un'applicazione web per l'indagine delle risposte a domande aperte provenienti da interviste telefoniche svolte mensilmente su un campione di utenza, sia privata che commerciale. L'applicazione, in linguaggio asp/VB Script, consente la ricerca avanzata di risposte tramite parole chiave e connettori logici, con la possibilità di filtrare la ricerca in base ad altre caratteristiche (ad esempio, tipologia di utenza o periodo di riferimento) e consente di ottenere statistiche sui profili degli utenti individuati in base all'interrogazione. Fino al 2006 si è occupato anche dell'aggiornamento dei dati mensili.

5)

Periodo: 2003 – 2004

Ditta o Ente: Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale della Facoltà di Economia di Sapienza - Università di Roma

Indirizzo: Via del Castro Laurenziano, 9 - 00161 Roma

Posizione: Collaboratore di ricerca

Mansioni svolte e risultati conseguiti:

Nel periodo 2003-2004 ha svolto attività di ricerca nell'ambito del progetto europeo NEMIS (<http://nemis.cti.gr/>), finanziato con fondi UE, in particolare in seno al working group 3 (WG3: Text mining applications, knowledge management & multilingual aspects). Il progetto in questione ha coinvolto gruppi di ricerca localizzati in Italia, Grecia, Spagna, Svizzera, Svezia, Lussemburgo e Francia, e si proponeva l'approfondimento e la divulgazione delle conoscenze acquisite nell'area del Text Mining. L'attività svolta in particolare ha riguardato la ricerca di settori di applicazione del Text Mining e l'individuazione delle tecniche e delle metodologie risolutive delle varie problematiche. Questo lavoro ha permesso, tra l'altro, di entrare in contatto con le principali aziende italiane del settore. Tra i settori di interesse del Text Mining, quelli che ricorrono più frequentemente sono la Pubblica Amministrazione e il mondo della ricerca, seguiti dalle società di telecomunicazioni e editoria, per loro natura interessate all'estrazione di informazione dall'enorme massa di informazione testuale di cui dispongono. Altri soggetti interessati al Text Mining sono gli istituti bancari, le società di servizi, insieme al settore sanitario e alle aziende farmaceutiche.

Le applicazioni individuate possono essere distinte in due campi principali, uno più strettamente riferito al campo del marketing, che comprende il *Customer Relationship Management*, la *Market Analysis* e le *Human Resources* e uno più tecnologico che comprende il *Technology Watch* e la *Patent Analysis*. Si tratta di applicazioni di Text Mining che rispondono alle esigenze specifiche del campo di applicazione e che quindi sono molto diverse tra loro, anche se tutte sono caratterizzate dall'esigenza di analizzare con sistematicità e in modo automatico grosse moli di dati non strutturati.

L'analisi delle esigenze di ogni applicazione ha permesso di desumere un quadro comune di strategia di analisi e trattamento del testo. I primi passi per qualsiasi operazione di Text Mining sono la selezione e acquisizione del testo, l'omogeneizzazione del formato utilizzato e la normalizzazione del testo. Vi sono poi le operazioni di riduzione o trasformazione del testo, come l'eliminazione delle parole vuote o la lemmatizzazione delle forme. Una volta terminata la fase preparatoria vi è la fase di analisi, contraddistinta dalla scelta dell'unità di analisi (parole o *multiword expressions*) e dall'estrazione di linguaggio significativo (parole chiave) che consente di caratterizzare il contenuto del testo. Alcune strategie di Text Mining prevedono poi il ricorso a motori di analisi semantica o concettuale che consentono l'indicizzazione del testo o la riconduzione del contenuto ai concetti più significativi. Risultati dell'analisi possono essere operazioni di classificazione automatica dei testi, raggruppamento dei testi in *cluster* omogenei per struttura o contenuto oppure estrazione di conoscenza sintetica dai documenti analizzati. I risultati del lavoro di ricerca sono confluiti nelle pubblicazioni 6, 8 e 10.

6)

Periodo: Maggio – Giugno 2003

Ditta o Ente: Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale della Facoltà di Economia di Sapienza - Università di Roma

Indirizzo: Via del Castro Laurenziano, 9 - 00161 Roma

Posizione: Collaboratore di ricerca

Mansioni svolte e risultati conseguiti:

Nel periodo Maggio-Giugno 2003 ha svolto attività di docenza, per 20 ore di lezione, relativa ad un corso intensivo di recupero e sostegno nell'ambito del corso di Statistica III.

7)

Periodo: Aprile – Dicembre 2002

Ditta o Ente: Istituto Regionale di Ricerca Educativa (IRRE) del Lazio

Indirizzo: Via Guidubaldo del Monte, 54 - 00197 Roma

Posizione: Collaboratore di ricerca

Mansioni svolte e risultati conseguiti:

Nel 2002 ha svolto attività di collaborazione presso l'Istituto Regionale di Ricerca Educativa (IRRE) del Lazio ai fini dell'attività di monitoraggio del progetto "Comunicazione ed Educazione Alimentare", promosso come attività progettuale nelle scuole dell'obbligo del Lazio. In particolare ha svolto: un'analisi di quadro generale delle scuole partecipanti al progetto, desunta dal database delle schede di adesione allo stesso; un'analisi del questionario proposto ai partecipanti ai corsi di aggiornamento relativi al progetto suddetto; una rilevazione dei contenuti di sei focus groups somministrati a quattro diversi tipi di soggetti coinvolti nell'attuazione del progetto e successiva analisi dei testi raccolti; la messa a punto di un questionario di monitoraggio sulle attività svolte nelle scuole ed elaborazioni statistiche dei dati così rilevati. I risultati delle analisi sono confluiti nella pubblicazione 11.

8)

Periodo: 2001 – 2003

Ditta o Ente: Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale della Facoltà di Economia di Sapienza - Università di Roma

Indirizzo: Via del Castro Laurenziano, 9 - 00161 Roma

Posizione: Collaboratore di ricerca

Mansioni svolte e risultati conseguiti:

Tra il 2001 ed il 2003 ha partecipato alla messa a punto e somministrazione di un questionario di valutazione della didattica universitaria (soli moduli di primo e secondo anno) presso la Facoltà di Economia di Sapienza - Università di Roma, attività svolta in seno all'Osservatorio Statistico della Facoltà stessa. Si è inoltre occupato delle analisi dei dati raccolti e i risultati sono stati pubblicati sulla pagina dell'Osservatorio Statistico di Facoltà <http://3w.eco.uniroma1.it/osservstat/>. I risultati della sola indagine relativa all'a.a. 2002/2003 sono presentati nella relazione tecnica.

Gli esiti delle varie rilevazioni sono stati utilizzati per ricalibrare opportunamente i programmi dei moduli di lezione coinvolti, favorendone l'armonizzazione. Inoltre, tramite metodi di analisi multidimensionale, è stato possibile cogliere le principali relazioni tra l'impegno profuso dagli studenti ed i risultati ottenuti in termini di esami superati.

9)

Periodo: 2001 -2002

Ditta o Ente: Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale della Facoltà di Economia di Sapienza - Università di Roma

Indirizzo: Via del Castro Laurenziano, 9 - 00161 Roma

Posizione: Collaboratore di ricerca

Mansioni svolte e risultati conseguiti:

Costruzione di un database di riferimento del linguaggio della stampa nazionale (quotidiani) relativo al periodo 1990-1999. Il lavoro è proseguito con l'analisi morfologica delle alterazioni e delle derivazioni linguistiche, con l'intento di arrivare al riconoscimento automatico di dette forme attraverso l'implementazione in software per analisi statistica dei dati testuali delle conoscenze acquisite. I testi raccolti per costruire il database sono stati anche analizzati con lo scopo di catturare le principali evoluzioni del linguaggio della stampa avvenute nella decade presa in considerazione. In quest'ottica, il sottoscritto ha sviluppato un indice di evoluzione temporale che permette di distinguere parole con trend di utilizzo crescente piuttosto che decrescente, nonché neologismi ed obsolescenze (si veda pubblicazione 12). Oltre al confronto di tipo diacronico, il linguaggio della stampa è stato comparato rispetto ad altre tipologie di registri (parlato, standard, formale), con lo scopo di identificarne le similarità e differenze di fondo. Il confronto è avvenuto sulla base di alcune caratteristiche generali, quali le incidenze delle diverse categorie grammaticali, dei tempi e modi verbali, dei generi, delle persone, ovvero sia l'impronta (*imprinting*) caratteristica del testo (si veda pubblicazione 7).

LINGUE STRANIERE

	Scolastico	Buono	Ottimo generico	Ottimo professionale
Inglese	X			
Francese				
Tedesco				
Altro.....				
Altro.....				

CONOSCENZE INFORMATICHE

Strumenti:

Microsoft Office (Word, Excel, Access, Power Point)

Tipo:

Pacchetto applicativo

Strumenti:

Internet explorer, Opera, Firefox, WS_FTP e client di posta elettronica

Tipo:

Browser, client ftp e client di posta elettronica

Strumenti:

Spad, R, SPSS e SAS Enterprise Miner

Curriculum vitae di Alessio Canzonetti

Tipo:

Software di analisi statistica

Strumenti:

Taltac, WordMapper, Coocs, TreeTagger, Lexico, Lexical Studio

Tipo:

Software di analisi automatica dei testi e statistica testuale

Strumenti:

Macromedia Dreamweaver, Microsoft Visual Studio (conoscenza codice html e Visual Basic)

Tipo:

Ambienti di sviluppo software e siti html

Strumenti:

TextPad, UltraEdit32, WinMerge

Tipo:

Editor di testo

TITOLI VARI E NOTE

Dal 2006 svolge attività di programmazione nell'ambito dello sviluppo del software per analisi dei dati testuali TALTAC (www.taltac.it). Il software, sviluppato in linguaggio Visual Basic, consente l'analisi automatica dei testi attraverso l'estrazione del vocabolario dei termini utilizzati e della loro frequenza, funzioni di normalizzazione del testo, confronto con lessici di frequenza, estrazione del linguaggio peculiare, analisi delle specificità, individuazione dei segmenti ripetuti, ricerche nei testi in base a regole, creazione di variabili desunte dal testo, calcolo di indici di dispersione ed uso dei termini, classificazione semantica, ricostruzione del testo completo di annotazioni generate in fase di analisi.

Il sottoscritto è a conoscenza che, ai sensi dell'art. 76 del D.P.R. del 28 dicembre 2000, n. 445, le dichiarazioni mendaci, la falsità negli atti e l'uso di atti falsi sono puniti ai sensi del codice penale e delle leggi speciali.

Firma

19-12-2008

Alessio Canzonetti