

Cod. SEL. 2CO/2012**Curriculum Vitae****Informazioni personali**

Nome / Cognome Alessio Canzonetti

Indirizzo

Telefono

Fax

E-mail

Cittadinanza

Data di nascita

Sesso Maschio

Istruzione e formazione

Data 11 luglio 2001

Titolo della qualifica rilasciata Laurea in Economia

Principali tematiche/competenze professionali possedute 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 - votazione: 104/110

Nome e tipo d'organizzazione erogatrice dell'istruzione e formazione Università degli Studi di Roma "La Sapienza"

Livello nella classificazione nazionale o internazionale ISCED 5

Data 25, 26 e 27 giugno 2007

Titolo della qualifica rilasciata Corso di formazione su software statistico – "Statistica multivariata con R"

Principali tematiche/competenze professionali possedute Il corso ha illustrato l'esecuzione di analisi multidimensionali attraverso il software R mediante richiami teorici, esempi ed applicazioni. In particolare sono stati affrontati le seguenti tecniche:

- Sintesi grafiche per dati multivariati
- Analisi in Componenti Principali (ACP)
- Analisi Fattoriale
- Analisi delle Corrispondenze
- Richiami di Regressione Lineare Multipla
- ACP e Regressione
- Partial Least Squares
- Scaling Multidimensionale
- Cluster Analysis (incl. Fuzzy Clustering e cenni a model-based clustering)

Nome e tipo d'organizzazione erogatrice dell'istruzione e formazione Facoltà di Scienze Statistiche – Sapienza Università di Roma

Data 26, 27 e 28 febbraio 2007

Titolo della qualifica rilasciata Corso di formazione su software statistico – "Corso introduttivo all'utilizzo di R"

Principali tematiche/competenze professionali possedute Il corso ha illustrato i principi generali dell'ambiente R e la sua sintassi attraverso esempi ed applicazioni. Le applicazioni sono state anche lo spunto per vedere in pratica l'utilità degli strumenti di statistica descrittiva (sintesi e rappresentazioni grafiche) ed inferenziale (intervalli di confidenze e test di ipotesi). In particolare sono stati affrontati i seguenti argomenti:

- Sintassi di base e strutture dati
- Importazione e pulizia di data-set
- Grafici
- Statistiche descrittive
- Elementi di statistica Inferenziale (intervalli di confidenza e principali test di ipotesi)
- Regressione

Nome e tipo d'organizzazione erogatrice dell'istruzione e formazione Facoltà di Scienze Statistiche – Sapienza Università di Roma

Esperienza professionale

	1
Date	Dal Marzo 2012
Nome e indirizzo del datore di lavoro	Consorzio MIPA - Consorzio per lo Sviluppo delle Metodologie e delle Innovazioni nelle Pubbliche Amministrazioni - Via Arenula 16, 00186 - Roma
Tipo di attività o settore	Il Consorzio promuove, progetta e svolge studi e ricerche sull'organizzazione e il funzionamento delle amministrazioni pubbliche centrali, regionali e locali. Sviluppa e diffonde metodologie per l'innovazione organizzativa e gestionale, la valutazione dell'azione amministrativa, il benchmarking e il miglioramento della qualità. Progetta e realizza azioni formative fondate sulle attività di ricerca.
Lavoro o posizione ricoperti	Collaboratore esterno di ricerca
Principali attività e responsabilità	<p><i>Progetto Roma Sad-Anr sperimentale</i></p> <p>L'Amministrazione di Roma Capitale, con apposita delibera di Giunta Comunale, ha avviato il "Progetto di sperimentazione di un nuovo modello di assistenza domiciliare per persone con disabilità e anziani". La sperimentazione mira ad individuare un nuovo criterio di compartecipazione alle spese del servizio da parte degli utenti basato sull'Isee (indicatore di situazione economica equivalente). In particolare, ha implicato una ridefinizione dell'Isee standard allo scopo di tenere in considerazione alcuni parametri ritenuti importanti per la formazione di un indicatore (detto Isee-Anr) in base al quale impostare i criteri di contribuzione ai servizi di cui i cittadini possono usufruire.</p> <p>La sperimentazione è stata portata avanti in quattro municipi (Municipio 2, 4, 8 e 13) ed ha riguardato 1.756 domande di richiesta di servizio in totale. In questo quadro, il sottoscritto ha curato le analisi quantitative in relazione alla situazione corrente della sperimentazione. Ha inoltre utilizzato i dati raccolti per proporre e confrontare lo scenario derivante da altre proposte alternative.</p> <p><i>Progetto Iseenet-Rsa e Iseenet-Pr</i></p> <p>La Regione Lazio, a partire dal 1 maggio 2007, ha introdotto l'utilizzo dell'Indicatore della situazione economica equivalente (Isee) per il calcolo della quota contributiva a carico degli assistiti nelle Residenze sanitarie assistenziali (Rsa). Più nello specifico, l'Amministrazione regionale, con la D.G.R. n. 98 del 20 febbraio 2007, ha rimodulato gli oneri della diaria giornaliera per le Rsa, suddividendoli al 50% fra il Fondo Sanitario Regionale (FSR) e gli assistiti, i quali possono beneficiare di un'eventuale contribuzione alla spesa da parte del Comune di residenza. La quota contributiva a carico dell'assistito viene determinata in funzione del valore Isee e di alcuni altri parametri, mentre il contributo a carico del Comune corrisponde all'importo che, sommato alla quota assegnata all'assistito, dà come risultato il 50% della tariffa complessiva da versare alla Rsa. Roma Capitale ha incaricato il Consorzio Mipa di sviluppare un apposito software (Iseenet Rsa) per gestire le domande di compartecipazione alla spesa e calcolare le quote contributive, affidandogli inoltre interventi di formazione rivolti al personale dei Municipi deputato agli inserimenti. Il Mipa ha svolto le suddette attività lavorando in stretta sinergia e collaborazione con l'ufficio responsabile del servizio (Dipartimento Promozione dei Servizi Sociali e della Salute - Direzione Coordinamento Servizi Sociali alla Persona e alle Famiglie e Promozione della Salute - U.O. Disabilità-Disagio Mentale e Promozione della Salute - Ufficio RSA).</p> <p>Il sottoscritto ha effettuato le analisi quantitative per l'anno 2011 relative al monitoraggio dello stato dell'arte in merito all'applicazione della normativa regionale da parte di Roma Capitale. Tale analisi va ad aggiornare i risultati delle elaborazioni svolte periodicamente dal Mipa, consentendo in tal modo di arricchire la serie storica già costruita.</p>

Analisi D'impatto Della Regolamentazione (Air): Modello Di Archiviazione Della Documentazione Amministrativa Di Roma Capitale

Progetto realizzato in collaborazione con il Nucleo Air del Comune di Roma Capitale e con l'Archivio Capitolino (Settembre 2011-Giugno 2012). Nel contesto generale del processo di "E-Gov" e dematerializzazione di Roma Capitale, l'obiettivo dell'Air è quello di definire/affinare il modello organizzativo e procedurale di archiviazione della documentazione amministrativa, con interventi di miglioramento del sistema di protocollo informatico. Lo studio si propone di valutare l'impatto interno all'amministrazione di tale innovazione organizzativo/gestionale. Oltre all'analisi della normativa e delle prassi vigenti in materia di classificazione dei documenti, ha previsto un articolato piano di consultazioni, condotto attraverso la realizzazione di focus group e questionari strutturati, nonché la richiesta di dati direttamente presso il Comune. Le risultanze emerse in questa fase, unitamente ai contributi provenienti da una serie di tavoli tecnici, hanno permesso di definire un sistema di classificazione dei documenti da implementare nell'applicativo di protocollazione attualmente in uso presso l'amministrazione capitolina. All'interno del progetto il sottoscritto ha condotto una serie di analisi quantitative volte, da un lato, alla documentazione dell'intensità e delle caratteristiche del fenomeno in esame (numero di documenti coinvolti nella revisione del processo, in totale, per tipologia e per struttura organizzativa, consistenza e distribuzione del personale addetto alla protocollazione ecc), dall'altro, si è attivato per individuare indicatori di efficienza e di standard dell'attuale processo gestionale sulla base dei dati disponibili. Ha quindi curato stesura del report finale in merito agli argomenti affrontati.

Linee metodologiche Regolamenti Air e Vir di attuazione della L. r. Puglia, 2/11/2011, n. 29

La L. r. 29/2011 introduce per la prima volta nella Regione Puglia strumenti normativi di analisi e controllo della qualità della regolazione, quali raccomandazioni al ricorso a testi unici quando possibile, alla chiarezza e alla semplicità nel drafting normativo e alle consultazioni con i soggetti destinatari dei provvedimenti.

La legge introduce quindi l'analisi (Air, art. 6) e la verifica (Vir, art. 8) di impatto della regolamentazione nel processo di normazione regionale, ribadendo la necessità di provvedere ad analisi tecnico-normativa (Atn, art. 7) e l'opportunità di prevedere in taluni casi clausole valutative (art. 9) all'interno degli articolati.

Il sottoscritto ha curato la stesura delle metodologie per la realizzazione dell'Air e della Vir nelle parti riguardanti la definizione degli obiettivi, i criteri di valutazione delle diverse opzioni di regolazione e le principali tecniche di consultazione.

Verifica di impatto della regolazione (VIR): Sistema di autorizzazione al funzionamento delle strutture e dei servizi sociali e relative forme di gestione.

Oggetto della verifica è il sistema di autorizzazione al funzionamento delle strutture e dei servizi sociali e delle relative forme di gestione nella Regione Puglia (introdotto con l.r. n. 19/2006 e Reg. n. 4/2007). Occorre verificare se la normativa sul sistema integrato ha trovato un riscontro certo ed effettivo sul piano applicativo, in termini di incremento di efficienza e di efficacia delle strutture preposte all'erogazione dei servizi. Pertanto, oggetto principale della valutazione sono le unità di offerta attualmente operative nell'erogazione dei servizi sociali ed assistenziali in Puglia, il rapporto con l'utenza, il comportamento delle amministrazioni pubbliche interessate.

Il sottoscritto ha partecipato alla revisione e all'integrazione dei risultati dell'analisi.

	2
Date	Settembre 2009 – Febbraio 2012
Nome e indirizzo del datore di lavoro	Istat - Via Cesare Balbo 16 00184 - Roma
Tipo di attività o settore	Trattamento e analisi di dati statistici
Lavoro o posizione ricoperti	Collaboratore esterno di ricerca
Principali attività e responsabilità	<p>Collaborazione con l'ISTAT presso il Dipartimento per le statistiche sociali ed ambientali - Direzione centrale delle statistiche socio-demografiche e ambientali – Servizio sanità, salute ed assistenza, in relazione all' "Indagine censuaria sugli interventi e i servizi sociali dei Comuni singoli e associati", che rileva informazioni sugli utenti e sulla spesa sostenuta dai Comuni per i servizi che erogano da soli e/o in associazione con altri, secondo un modello di organizzazione che può variare non soltanto tra le regioni, ma anche all'interno della stessa regione.</p> <p>L'unità di rilevazione dell'indagine è costituita dai Comuni singoli e dalle loro associazioni quali comprensori e consorzi, oltre che da Comunità montane e da distretti socio-sanitari/ASL che affiancano i Comuni singoli e talvolta li sostituiscono.</p> <p>L'indagine è frutto della collaborazione con più enti del governo centrale e locale, interessati alla produzione e all'utilizzo della base dati a cui la rilevazione è finalizzata. Oltre al coinvolgimento della Ragioneria generale dello Stato (RGS) e, quindi, del Ministero dell'economia e delle finanze (MEF), hanno partecipato direttamente alla rilevazione il Ministero del lavoro e delle politiche sociali e il Centro Interregionale per il Sistema Informatico e Statistico (CISIS) con alcuni enti locali.</p> <p>Il sottoscritto si è occupato principalmente del controllo della congruità dei dati comunicati dagli enti censiti attraverso procedure, in codice SAS ed R, prodotte in prima persona, allo scopo di individuare duplicazioni, errori di inserimento, misinterpretazioni delle grandezze richieste nel questionario (<i>on-line</i>) da parte dei fornitori dei dati, e produrre quindi richieste di eventuale rettifica del dato comunicato, con conseguente integrazione delle stesse. Ha partecipato inoltre alla definizione dell'intero flusso di raccolta e trattamento dei dati e ai processi di stima dei dati non pervenuti. Ha provveduto alla stesura di procedure per la produzione degli output in forma gabellare, livello di provincia, regione e ripartizione territoriale, e al calcolo dei principali indicatori di sintesi circa lo stato dei servizi. Ha contribuito all'impostazione di una sperimentazione riguardante la rilevazione degli asili nido, che estenderebbe l'osservazione anche agli istituti privati. Il tutto è avvenuto operando nell'ottica del raggiungimento della massima standardizzazione, ripetibilità ed automazione dell'intero processo.</p>
	3
Date	Febbraio 2009 – Luglio 2009
Nome e indirizzo del datore di lavoro	Centro Interdipartimentale di Biostatistica e Bioinformatica – Università di Roma Tor Vergata - Via Orazio Raimondo 18, 00173, Roma
Tipo di attività o settore	Ricerca universitaria
Lavoro o posizione ricoperti	Assegnista di ricerca
Principali attività e responsabilità	<p>Il titolo della ricerca è "Analisi di dati relativi a fenomeni droga-correlati, costruzione di modelli interpretativi, costruzione e stima di indicatori". Sulla base di dati forniti dall'Autorità Giudiziaria e dai centri di recupero, sono stati applicati metodi di stima con riguardo al numero dei tossicodipendenti in Italia e al giro d'affari generato dal mercato illegale degli stupefacenti, in totale e per tipo di sostanza prevalentemente consumata, proponendo una stima delle dimensioni del mercato delle sostanze stupefacenti e della sua incidenza sul prodotto interno lordo nazionale, attraverso un approccio lato domanda e lato offerta. Sono stati inoltre posti in evidenza i limiti dell'approccio, dovuti essenzialmente alla qualità dei dati disponibili. La raccolta da parte dei fornitori non ha infatti finalità statistiche, ma segue semplicemente esigenze di prassi e protocollo interne che vanno spesso a determinare una non uniformità dei dati stessi.</p> <p>Ho inoltre collaborato all'analisi dei dati ottenuti attraverso un questionario <i>on-line</i> proposto agli studenti di un campione di istituti scolastici italiani, occupandosi in particolare del confronto con le evidenze emerse da un'analoga indagine operata negli Stati Uniti. Questa viene portata avanti fin dal 1975, e possiede un <i>datawarehouse</i> estremamente ricco ed articolato, accessibile <i>on-line</i> attraverso un'interfaccia che offre peraltro la possibilità di analisi dei dati e la possibilità di scaricare i microdati e la relativa, dettagliatissima, documentazione. Attraverso l'utilizzo di questo <i>datawarehouse</i>, si sono potuti desumere informazioni relative alle esperienze maturate in questi anni hanno che offerto chiavi di lettura e approcci di sicura utilità anche per la realtà italiana.</p>

	4
Date	2007 - 2012
Nome e indirizzo del datore di lavoro	
Tipo di attività o settore	Trattamento e analisi di dati statistici
Lavoro o posizione ricoperti	Collaboratore di ricerca
Principali attività e responsabilità	Dal 2007 fa parte di un gruppo di ricerca, guidato dal Prof. Alberto Zuliani, che porta avanti un progetto di analisi statistica delle sentenze della Corte dei Conti in materia di responsabilità amministrativa. In particolare il sottoscritto si occupa della gestione tecnica dei materiali di analisi, del coordinamento delle operazioni di inserimento dati, della produzione delle interfacce e delle strutture di raccolta dei dati, nonché della produzione delle elaborazioni statistiche. Attualmente la ricerca si è allargata anche ai provvedimenti del Consiglio di Stato, per cui il sottoscritto sta producendo il relativo database di raccolta e il sistema di <i>data entry</i> per l'inserimento guidato dei dati. I risultati raggiunti al momento sono stati presentati nella pubblicazione 2, mentre è in corso di pubblicazione un'ulteriore analisi basata su un campione di sentenze di primo grado in materia di responsabilità amministrativa.
	5
Date	Aprile 2006 – Marzo 2008
Nome e indirizzo del datore di lavoro	Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale della Facoltà di Economia di Sapienza - Università di Roma Via del Castro Laurenziano, 9 - 00161 Roma
Tipo di attività o settore	Ricerca universitaria
Lavoro o posizione ricoperti	Assegnista di ricerca
Principali attività e responsabilità	<p>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, 3 e 4). Le conoscenze acquisite in fase di ricerca sono state implementate e testate attraverso il software TALTAC.</p> <p>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), si disponeva anche di informazioni strutturate. L'analisi ha quindi 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 5).</p> <p>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 svolto 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, <i>cluster analysis</i> 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. Sono inoltre state impartite nozioni di <i>data warehousing</i>, <i>data mining</i> e <i>On-Line Analytical Processing</i>.</p>

	6
Date	2006 - 2012
Nome e indirizzo del datore di lavoro	CITICORD di Sapienza – Università di Roma P.le Aldo Moro, 5 – Roma e Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale della Facoltà di Economia di Sapienza - Università di Roma Via del Castro Laurenziano, 9 - 00161 Roma
Tipo di attività o settore	Formazione
Lavoro o posizione ricoperti	Docente
Principali attività e responsabilità	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 22, sia di livello introduttivo al software, sia di livello avanzato, ed hanno coinvolto circa 150 partecipanti in totale provenienti da oltre 40 dipartimenti in Italia.
	7
Date	15 Luglio 2004 – 15 Settembre 2004
Nome e indirizzo del datore di lavoro	Dipartimento di Scienze Umane dell'Università per Stranieri di Siena Piazza Carlo Rosselli 27/28- 53100 Siena
Tipo di attività o settore	Ricerca universitaria
Lavoro o posizione ricoperti	Collaboratore di ricerca
Principali attività e responsabilità	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.
	8
Date	2004 –2006
Nome e indirizzo del datore di lavoro	BSD s.r.l. Via Trieste, 32 - 35121 Padova
Tipo di attività o settore	Information Technology
Lavoro o posizione ricoperti	Collaboratore esterno
Principali attività e responsabilità	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 un datawarehouse di interviste telefoniche, svolte mensilmente da una società demoscopia su mandato della committenza, 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, profili individuati in base alle variabili raccolte nel corso dell'intervista (età, sesso, condizione socio-professionale, numero componenti della famiglia, possesso di collegamento ad internet ecc.). Fino al 2006 si è occupato anche dell'aggiornamento trimestrale dei dati mensili.

	9
Date	2003 – 2004
Nome e indirizzo del datore di lavoro	Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale della Facoltà di Economia di Sapienza - Università di Roma Via del Castro Laurenziano, 9 - 00161 Roma
Tipo di attività o settore	Ricerca universitaria
Lavoro o posizione ricoperti	Collaboratore di ricerca
Principali attività e responsabilità	<p>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.</p> <p>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.</p> <p>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 7, 9 e 11.</p>
	10
Date	Maggio – Giugno 2002
Nome e indirizzo del datore di lavoro	Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale della Facoltà di Economia di Sapienza - Università di Roma Via del Castro Laurenziano, 9 - 00161 Roma
Tipo di attività o settore	Istruzione universitaria
Lavoro o posizione ricoperti	Docente
Principali attività e responsabilità	Nel periodo Maggio-Giugno 2002 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. Il contenuto del corso prevedeva il trattamento delle tecniche di analisi multidimensionale dei dati.

	11
Date	Aprile – Dicembre 2002
Nome e indirizzo del datore di lavoro	Istituto Regionale di Ricerca Educativa (IRRE) del Lazio Via Guidubaldo del Monte, 54 - 00197 Roma
Tipo di attività o settore	Rilevazione ed analisi di dati statistici
Lavoro o posizione ricoperti	Collaboratore di ricerca
Principali attività e responsabilità	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 12.
	12
Date	2001 –2003
Nome e indirizzo del datore di lavoro	Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale della Facoltà di Economia di Sapienza - Università di Roma Via del Castro Laurenziano, 9 - 00161 Roma
Tipo di attività o settore	Rilevazione ed analisi di dati statistici
Lavoro o posizione ricoperti	Collaboratore di ricerca
Principali attività e responsabilità	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. La rilevazione ha avuto cadenza semestrale, ed ha coinvolto un numero di studenti via via crescente (dai 600 dell'indagine pilota fino ai quasi 1200 dell'ultima rilevazione), con l'intento di coprire al meglio la tipologia degli studenti frequentanti. I dati raccolti sono andati a costituire una preziosa fonte informativa in serie storica che hanno permesso il monitoraggio della situazione in merito all'attuazione del nuovo ordinamento didattico. 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, tramite l'implementazione di indicatori di performance è stato possibile cogliere le principali relazioni tra l'impegno profuso dagli studenti ed i risultati ottenuti.
	13
Date	2001 –2002
Nome e indirizzo del datore di lavoro	Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale della Facoltà di Economia di Sapienza - Università di Roma Via del Castro Laurenziano, 9 - 00161 Roma
Tipo di attività o settore	Ricerca universitaria
Lavoro o posizione ricoperti	Collaboratore di ricerca
Principali attività e responsabilità	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 un 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 13). 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 la impronta (<i>imprinting</i>) caratteristica del testo (si veda pubblicazione 8).

Capacità e competenze personali

Negli ultimi anni ha sviluppato competenze e interesse nell'ambito della raccolta e del trattamento dei dati statistici, curando in particolar modo gli aspetti riguardanti la corretta integrazione tra le varie fasi del processo e tra i vari sistemi in esso eventualmente conviventi. Ha maturato esperienza anche relativamente al trattamento delle mancate risposte parziali e totali e all'imputazione dei dati mancanti.

Inoltre, è attivamente coinvolto nell'ambito dell'analisi statistica dei dati testuali e del Text Mining. In questo quadro si è occupato di trattamento del linguaggio naturale (NLP - *Natural Language Processing*), analisi multidimensionale dei dati, trattamento dati, tecniche di classificazione e *cluster analysis*. Tale capacità è stata utilizzata anche in ambito di codifica dei quesiti aperti.

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.

Per la persistenza dei dati il software si appoggia ad un motore di database esterno, per cui il sottoscritto possiede ottima padronanza del linguaggio SQL e della gestione dei database relazionali in generale, non solo per quanto riguarda la definizione della struttura entità-relazioni di un database, ma anche per ciò che concerne il processo di *data-entry* (realizzazione interfaccia grafica e implementazione dei controlli in fase di inserimento dati). Queste conoscenze sono state spesso utilizzate anche in progetti estranei allo sviluppo del software, nell'ambito di indagini più strettamente statistiche, mediante le quali ha sviluppato capacità di disegno e progettazione delle medesime.

In vista di una riprogettazione del software ha cominciato ad approfondire la conoscenza di altri linguaggi di programmazione di moderna estrazione, quali Ruby e Python.

Madrelingua(e)

Altra(e) lingua(e)

Autovalutazione

Livello europeo

Inglese

Inglese

Comprensione		Parlato		Scritto
Ascolto	Lettura	Interazione orale	Produzione orale	
B1	B2	A2	A2	B1

Conoscenze informatiche

Produttività personale

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

LibreOffice (Writer, Calc, Impress)

Software di analisi statistica

SAS e SAS Enterprise Miner, R, SPSS, Spad

Software di analisi automatica dei testi e statistica testuale

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

Ambienti di sviluppo software e siti html

Eclipse, Microsoft Visual Studio, NetBeans, SciTe, Macromedia Dreamweaver.

Conoscenza codice Visual Basic, asp/VBScript, Ruby, Python, R, html e xml.

Browser, client ftp e client di posta elettronica

Firefox, Chrome, Internet explorer, WS_FTP, Thunderbird, Outlook Express

Sistemi operativi:

Ubuntu/Linux, Windows

Pubblicazioni

- 1
- Autori: Canzonetti Alessio
- Titolo: Semantic classification and cooccurrences: a method for the rules production for the information extraction from textual data (versione estesa), 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", pp. 209-216, Springer-Verlag, Berlin-Heidelberg.
- Data pubblicazione: 2010
- 2
- Autori: Zuliani A., Aurisicchio G., Canzonetti A., De Benedetto M., Liverani A., Menichino P., Rispoli L., Salvi S.
- Titolo: Un'analisi statistica delle sentenze della Corte dei Conti: prime evidenze, in *Rivista trimestrale di diritto pubblico*, Anno LIX Fasc. 3, pp. 673-706, Giuffrè Editore, Milano.
- Data pubblicazione: 2009
- 3
- 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.
- Data pubblicazione: 2008
- 4
- 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 pubblicazione: 2007
- 5
- Autori: Canzonetti A., Misuraca M., Romano M.C.
- Titolo: Analyzing the language of everyday life: how Textual Statistic can support Time Use Surveys. The Italian Experience, in *Società Italiana di Statistica - Atti della XLIII Riunione Scientifica*, pp. 69-80, Torino.
- Data pubblicazione: 2006
- 6
- Autori: Canzonetti A., Capo F.M., Ruocco V.
- Titolo: 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 pubblicazione: 2006
- 7
- 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 pubblicazione: 2005

- 8
 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.
 Data pubblicazione: 2005
- 9
 Autori: Bolasco S., Canzonetti A., Capo F.M.
 Titolo: *Text Mining: uno strumento strategico per imprese e istituzioni*, CISU, Roma.
 Data pubblicazione: 2005
- 10
 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 pubblicazione: 2004
- 11
 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 pubblicazione: 2004
- 12
 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 pubblicazione: 2004
- 13
 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 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: 2005

Partecipazioni a convegni e seminari

Settembre 2009: Workshop on *Illicit drug market and its possible regulatory body*, Presentazione dei contributi "Il mercato: quantità, consumi, valori" e "Esperienze internazionali: Monitoring the future - Indagine su abitudini e stili di vita degli studenti USA", Facoltà di Economia, Sapienza – Università di Roma.

Marzo 2008: *JADT 2008 Journées Internationales d'Analyse Statistique des Données Textuelles*, presentazione del paper "Information retrieval e analisi delle cooccorrenze per l'estrazione di informazione specifica da documentazione giuridica", ENS-LSH Lyon.

Settembre 2007: CLADAG 2007, presentazione del paper "Semantic classification and cooccurrences: method for the rules production for the information extraction from textual data", Università di Macerata.

Giugno 2006: XLIII Riunione Scientifica della Società Italiana di Statistica, presentazione del paper "Analyzing the language of everyday life: how Textual Statistic can support Time Use Surveys. The Italian Experience", Università degli Studi di Torino.

Aprile 2006: *JADT 2006 Journées Internationales d'Analyse Statistique des Données Textuelles*, presentazione del paper "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", Besançon,.

Ottobre 2004: Partecipazione alla Nemis Final Conference "Knowledge Mining" – Atene, organizzato da Research Academic Computer Technology Institute.

Gennaio 2004: Organizzazione e partecipazione al 2° Nemis Annual Conference "Text Mining for Business Intelligence", Facoltà di Economia di Sapienza - Università di Roma.

Settembre 2003: CLADAG 2003, presentazione del paper "Some insights into the evolution of 1990s' standard Italian using Text Mining techniques and automatic categorisation", Università di Bologna.

Giugno 2003: Riunione Scientifica della Società Italiana di Statistica, presentazione del paper "Text mining as key to reading the Nineties' language in the daily "La Repubblica"", Università di Napoli "Federico II".

Marzo 2003: Partecipazione al seminario "Modelli statistici per il web mining", tenuto da Paolo Giudici presso l'Istituto di Metodi Quantitativi dell'Università "Bocconi" di Milano.

Ulteriori informazioni

Allegati documento di identità in corso di validità

Il sottoscritto dichiara di essere informato, ai sensi e per gli effetti di cui all'art. 13 del D. Lgs. n. 196/2003 che i dati personali raccolti saranno trattati, anche con strumenti informatici, esclusivamente nell'ambito del procedimento per il quale la presente dichiarazione viene resa.

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

Alla presente dichiarazione allega copia fotostatica leggibile, fronte-retro, non autenticata di un documento di identità in corso di validità.

Luogo e data
Orbetello, 6 dicembre 2012

Firma

