

## ***Curriculum vitae* di SERGIO BOLASCO**

sergio.bolasco@uniroma1.it

Dip. Metodi e Modelli per l'Economia, il Territorio e la Finanza (MEMOTEF)

Via del Castro Laurenziano 9 – 00161 – Roma (Italy)

Attualmente docente a contratto di "*Metodi esplorativi per l'analisi dei dati e laboratorio di data mining*" presso la Facoltà di Economia dell'Università di Roma "La Sapienza"

Già docente, nel Master di "Fonti Strumenti e Metodi per la Ricerca sociale" (FSeM) presso la Facoltà di Statistica dell'Università degli studi di Roma "La Sapienza", di un corso di "*Text mining e analisi dei dati testuali*" (fino al 2013).

### CURRICULUM ESSENZIALE

Laureato in Scienze Statistiche e Demografiche (1971).

Già professore ordinario di Statistica presso la Facoltà di Economia dell'Università di Roma "La Sapienza", afferente al Dipartimento di Metodi e Modelli per l'Economia, il Territorio e la Finanza (MEMOTEF), collocato a riposo dal 1 marzo 2012.

Nell'ultimo a.a. 2010-2011, titolare degli insegnamenti di "*Metodi esplorativi per l'analisi dei dati e laboratorio di data mining*", di "*Strumenti statistici per l'analisi quantitativa*" e di "*Metodi e tecniche per l'analisi dei dati*".

Nel periodo 2003-2009 nel corso di laurea specialistica "Metodi quantitativi per l'economia e l'impresa (METEI) presso la Facoltà di economia della SAPIENZA, di cui è stato fra i promotori e coordinatore, ha tenuto i corsi di "*Laboratorio avanzato di analisi dei dati e data mining*" e di "*Casi di studio in Data mining e Text mining*".

Già professore associato nei corsi di "*Statistica*" e di "*Matematica per le scienze sociali*" presso l'Università di Salerno, nel corso di laurea in Sociologia (1987-1995).

Già ricercatore presso la Facoltà di Scienze Statistiche, Dip. di Statistica, Probabilità e Statistiche Applicate dell'Università di Roma "La Sapienza" (1980-1987).

Professore incaricato presso la Facoltà di Scienze Sociali della Pontificia Università Gregoriana dal 1973 al 2003. In questi anni ha tenuto corsi di Matematica, di Statistica, di Analisi quantitativa e qualitativa, di Informatica e di Metodi di analisi automatica dei testi.

Dal 1978 è membro della Società Italiana di Statistica; è stato redattore della rivista internazionale METRON; dal 1993 al 1997 è stato rappresentante per l'Italia nel consiglio dell'International Federation of Classification Societies (IFCS).

Dal 1993 è membro del comitato scientifico internazionale del JADT (*Journées internationales d'Analyse statistique des Données Textuelles*), nel cui ambito ha organizzato a Roma, dapprima nel dicembre 1995 le Terze Giornate Internazionali JADT2005 e successivamente nel 2010 la loro decima edizione (<http://JADT2010.uniroma1.it>).

Nell'arco degli ultimi venti anni ha promosso, partecipato e diretto numerose ricerche finanziate di interesse nazionale e internazionale sui temi dell'analisi dei dati e sul *text mining* con particolare riferimento alle tecniche esplorative di tipo multidimensionale e ai metodi di analisi automatica dei testi. Fra le altre, ha coordinato l'unità operativa italiana nel progetto europeo NEMIS (*Network of Excellence in Text Mining and its Applications in Statistics*) negli anni dal 2002 al 2004 (<http://nemis.cti.gr>).

Ha svolto periodi di studio e ricerca all'estero nelle Università di Parigi, di Montpellier e di Lille in Francia, di Dubrovnic in Jugoslavia e di Trier in Germania.

Ha svolto seminari e scuole estive in varie Università italiane ed estere, fra cui: Besançon, Leon, Lille, Montreal, Paris, Toulouse, Bologna, Bolzano, Firenze, Modena, Napoli, Padova, Parma, Pavia, Perugia, Pisa, Pollenzo, Roma1, Roma2, Roma3, Salerno, Siena, Torino, Urbino, Viterbo.

Dal 2005 al 2009, in qualità di esperto straniero è stato membro del Conseil Scientifique de l'UMS 2913 Maison de Science de l'Homme di Besançon.

Dal 2007 al 2014 ha tenuto 27 corsi di formazione (3 gg) sull' "Analisi automatica dei testi" alla SAPIENZA Università di Roma, presso il CITICORD e presso il Dipartimento MEMOTEF.

Nel 2011, 2012 e 2013 ha organizzato tre edizioni di una Scuola internazionale sul tema "Introduzione all'analisi automatica dei testi e strumenti di text mining" di tipo interdisciplinare, che ha visto la partecipazione di linguisti, statistici, informatici, sociologi, psicologi e scienziati della comunicazione.

E' membro dell' Editorial Board delle riviste:

*International Review of Sociology*, edita da Routledge, Taylor and Francis Group  
(<http://www.tandfonline.com/action/aboutThisJournal?show=editorialBoard&journalCode=cirs20>)

*Lexicometrica*, rivista on-line dell'Università di Paris 3, fondata da André Salem,  
(<http://lexicometrica.univ-paris3.fr/comite.html>), .

*Linguisticae Investigationes*, Revue internationale de linguistique générale et française, fondata da Maurice Gross (<http://infolingua.univ-mlv.fr/LI/LINGVISTICAEINVESTIGATIONES.html>),

## COLLABORAZIONI CON ENTI NON UNIVERSITARI E ATTIVITÀ PROFESSIONALI

Dal 1975, ha svolto attività di promozione e/o direzione di numerose ricerche con funzioni di coordinamento scientifico, in Enti e Istituti di ricerca d'interesse nazionale o in Aziende, fra le quali: ABI - Associazione Bancaria Italiana; AGCM - Autorità Garante della Concorrenza e del Mercato; CENSIS - Centro Studi Investimenti Sociali; Culture s.r.l.; Elettra 2000; Fivol - Federazione Italiana per il Volontariato; Formstat; GEPI s.p.a.; Grandi Numeri; Isfort - Istituto superiore di formazione e ricerca per i trasporti; IRIT - Istituto di Ricerca per l'Industria e il Territorio; IRRE del Lazio; Istituto Tagliacarne; Federmeccanica; MIPA - Consorzio per lo sviluppo delle Metodologie e delle Innovazioni nella Pubblica Amministrazione; Progetto Europa s.r.l.; Provincia di Trento; RAI Radio Televisione Italiana; Regione Lazio; Regione Puglia; Regione Autonoma Valle d'Aosta; SATIN s.r.l. - Sviluppo Attività Industriali; Scuola Superiore della Pubblica Amministrazione; Seimart (Torino); SOMEA - Società per la Matematica e l'Economia Applicate; Telecom Italia.

Ha collaborato con i seguenti Editori: Aracne, Carocci Editore, Cisu, Editori Riuniti, Electa Editrice, Eri - Edizioni RAI, Franco Angeli, Il Mulino, LED Edizioni Universitarie, Manifesto Libri, Marzorati Editore, North Holland, Nuova Italia Scientifica, Sage, Casa Editrice Università "la Sapienza", Springer-Verlag.

Ha svolto attività di referee per: CLADAG, COMPSTAT, IFCS, JADT, LED, Lexicometrica, Mots, Rassegna Italiana di Psicologia, Statistica Applicata - Italian Journal of Applied Statistics, SIS – Società Italiana di Statistica, Springer.

## SETTORI DI INTERESSE SCIENTIFICO

Metodi e tecniche di Text Mining; metodologie per il trattamento automatico dei dati testuali; analisi delle domande aperte in survey qualitative, focus-group, interviste non direttive, messaggi/forum in Internet; analisi del contenuto e del discorso. Trattamento del linguaggio naturale; costruzione di lessici di frequenza, di risorse statistico-linguistiche e di grammatiche locali. Software per l'analisi automatica dei testi e il text mining.

Analisi dei dati: tecniche statistiche multidimensionali, analisi fattoriali, metodi di classificazione automatica e di segmentazione, metodi multiway. Analisi socio-economiche sul territorio. Studi di customer satisfaction. Semiologia grafica e trattamento grafico dell'informazione.

Nel 2000 ha ideato e curato lo sviluppo del software Taltac, piattaforma per il Trattamento Automatico Lessicale e Testuale per l'Analisi del Contenuto di un Corpus, diffuso in Italia e all'estero in oltre 130 fra Dipartimenti universitari e Istituti di ricerca. Questo software è stato ampliato nel tempo dalla versione 1 del 2000-2005, alla versione 2 rilasciata dal 2005 fino al 2014, grazie al contributo di dottorandi, assegnisti oltre che di tecnici informatici esterni al Dipartimento. Attualmente coordina lo sviluppo della versione TaLTaC3, utilizzabile su differenti sistemi operativi e multiprocessore, in via di rilascio nella versione beta.

## INTERVENTI A CONVEGNI INTERNAZIONALI PIÙ RECENTI

- Giugno 2012 – Vers la version 3.0 de Taltac. Demo Logiciel, JADT2012, Université de Liège.
- September 2011 – “A Two-Level Fuzzy Classification between Different Profiles of the Gastronomic Lexicon”, CLADAG, Pavia
- January 2011 – Metodi per l'estrazione di informazione: il peso delle parole e sistemi di regole, Scuola internazionale di *Introduzione all'analisi automatica dei testi e ai metodi e modelli di text mining* (corso ADT-TM)
- 18-20 giugno 2009 – «*Logique de l'analyse automatique d'un corpus*» MISAT - Méthodes Informatiques et Statistiques en Analyse de Textes, Ecole Thématique du CNRS-Institut SHS Besançon (<http://laseldi.univ-fcomte.fr/ecole/index.php>)
- October 2008 – “TaLTaC2: funzionalità lessicali e testuali per l'analisi automatica di un corpus” Convegno su “*Corpus Linguistics in Practical Applications*” - Università degli Studi della Tuscia – Viterbo
- January 2004 - "Text mining for Business Intelligence" - Nemis Annual Conference at the Faculty of Economics of the University of Rome "La Sapienza"
- December 2003 - Workshop on "Application of textual data analysis" at the Faculty of Statistics of the University of Rome "La Sapienza"
- January 2002 - "Taltac software Workshop" at the Faculty of Economics of the University of Rome "La Sapienza"

## CONFERENZE e RELAZIONI SUL TEXT MINING:

- gennaio 2001, Conferenza al Laboratoire d'Automatique et Documentaire Linguistique, Univ. de Paris VII, su "L'exploitation d'un dictionnaire de fréquence dans l'analyse automatique des textes"
- dicembre 2001, Conferenza su "*Statistical analysis of texts and political lexicon*", Dipartimento di Studi Storici e Politici, Univ. di Padova
- novembre 2004, Censis - Conferenza su "*Text mining e analisi testuale*"
- dicembre 2004, ABI - Workshop su "*Il Text Mining in banca: una nuova sfida per semplificare il flusso di contatti con il cliente*", 3<sup>a</sup> Conferenza ABI - CRM2004
- aprile 2006 - Keynote speaker al XIV Congresso GISCEL con la relazione "*Corpora e liste di frequenza d'uso: criteri e tecniche per l'analisi automatica dei testi*" – Università di Siena,
- agosto 2006 - scuola estiva ATO: *Quelques expériences de Analyse de Données Textuelles*. - UQAM, Università de Montreal (Canada)
- giugno 2007 - Workshop on "*Recent developments of Taltac*", at the Dipartimento di Studi Geoeconomici Linguistici Statistici Storici per l'Analisi regionale, SAPIENZA University of Rome.
- settembre 2007 - "*Automatic Dictionary and Rule-Based Systems for Extracting Information from Text*" al 6<sup>a</sup> Scientific Meeting of the Classification and Data Analysis Group of the Italian Statistical Society, CLADAG 2007, Università di Macerata.
- 10-11 marzo 2008 - Seminario su "*Exploitation des ressources statistique-linguistiques dans l'analyse d'un corpus à l'aide du logiciel Taltac2*" alla Maison de Science de l'Homme di Besançon
- 5-6 maggio 2008 – "*Dal lessico delle guide, i tipi della ristorazione in Italia*", Alma Graduate Scholl, Università di Bologna - GamberoRosso
- 12 ottobre 2008 - "*TaLTaC2: funzionalità lessicali e testuali per l'analisi automatica di un corpus*" Seminario su "Corpus Linguistics in Practical Applications" - Università della Tuscia (Viterbo).
- 12 dicembre 2008 "*Alcuni fondamenti dell'analisi dei dati testuali secondo un approccio metrico. Esempi e tipi di applicazioni*" - Seminari GIAT - Gruppo interdisciplinare per l'analisi dei testi Università degli Studi di Padova
- 27 aprile 2009 "*Due esempi di information mining elaborati da un punto di vista metrico con il software TaLTaC*" Seminari Fondazione Bordoni, FUB, Roma
- aprile-giugno 2009, corso "*Taltac in MADIT*", Facoltà di Psicologia, Università di Padova
- 6 ottobre 2010 – "*Presente e futuro del Text Mining*" – SAS Campus 2010, Facoltà di Economia, Università degli studi di Bologna
- 13 ottobre 2010 – "*La rappresentazione dei testi con l'aiuto della statistica*" Seminario LOL alle "Vetriere", Dipartimento di Studi Filologici, Linguistici e Letterari, SAPIENZA - Università di Roma
- 17 gennaio 2011 – "*Introduzione all'analisi dei dati testuali e ai metodi di text mining*", Istituto Superiore S. Anna, Pisa
- 18 marzo 2011 – "*Il lessico della critica gastronomica. Un'analisi basata su tecniche di text mining*" Facoltà di Scienze Gastronomiche – Pollenzo
- 29 novembre 2012 - "*L'analisi automatica dei testi. relazione introduttiva al IV Forum "Analisi Qualitativa"*, Fac. di Scienze dell'Educazione, Università di Roma III

## ***Elenco pubblicazioni più recenti di SERGIO BOLASCO***

- 25 Bolasco S. (2013). *L'analisi automatica dei testi. Fare ricerca con il text mining*. Carocci, Roma, pp. 410.
- 24 Bolasco S., Pavone P. (2012b). *How the text mining measures complex phenomena in official statistics*, Proceedings of the XLVI Scientific Meeting of the Italian Statistical Society, Roma, 20-22 giugno 2012
- 23 Bolasco S. (2012a). Introduction to the Automatic Analysis of Textual Data via a Case Study. in *Statistica Applicata - Italian Journal of Applied Statistics* (Special Issue: Advances in Textual Data Analysis and Text Mining), vol. 22(1), pp. 5-19.
- 22 Escoubas-Benveniste M.P., Floquet O., Bolasco S. (2012). Contribution empirique à l'étude du gérondif et du participe présent en français parlé et écrit, in *JADT 2012 : 11es Journées internationales d'Analyse statistique des Données Textuelles*, Liegi
- 21 Bolasco S. (2011b). Differenti profili del lessico gastronomico e della ristorazione italiana, in M. Bolasco e M. Trabucco, *Cronache golose. Vita e storie di cuochi italiani*, Slow Food Editore, Bra, pp. 253-266. (La Commissione del Bancarella Cucina ha assegnato il Premio "Angelo Paracucchi")
- 20 Bolasco S. (2011a). Taltac, una piattaforma per l'analisi automatica dei testi, in K. Krippendorff e G. La Rocca (eds.) *Atti del convegno internazionale "RiQGioS 2011" (Ricerca Qualitativa e Giovani studiosi)*, Università "Kore", Enna 1-3 settembre 2011
- 19 Bolasco S., Pavone P. (2011). A Two-Level Fuzzy Classification between Different Profiles of the Gastronomic Lexicon, in P. Cerchiello e C. Tarantola (eds.) *CLADAG 2011 Book of Abstracts*, 8th Scientific Meeting of the CLAssification and Data Analysis Group of the Italian Statistical Society, University of Pavia, September 7-9, 2011. Pavia University Press Editoria scientifica
- 18 Bolasco S. (2010d). Taltac2.10 Sviluppo, esperienze ed elementi essenziali di analisi automatica dei testi, LED, Milano, pp. 55. (<http://www.ledonline.it/Taltac/allegati/459-7-TALTAC-COL.pdf>)
- 17 Bolasco S., Chiari I., Giuliano L. (eds.) (2010c). *JADT2010. Statistical Analysis of Textual Data*, Proceedings of 10th International Conference JADT, LED, Milano 2 voll., pp. 1330.
- 16 Bolasco S. (2010b). Il riconoscimento automatico di locuzioni verbali con l'ausilio del software Taltac2, *Rassegna Italiana di Linguistica Applicata*, XLII 1-2 : 39-56. (ISSN 0033-9725).
- 15 Bolasco S., Pavone P. (2010). Automatic Dictionary and Rule-Based Systems for Extracting Information from Text, in Francesco Palumbo, Carlo Natale Lauro, Michael Greenacre (eds.) "Data Analysis and Classification" Proceedings of the 6th Conference of the Classification and Data Analysis Group of the Società Italiana di Statistica, Springer, Berlin-Heidelberg. pp. 189-198.
- 14 Bolasco S., Pavone P. (2008b). *Multi-class categorization based on cluster analysis and TFIDF*, in S. Heiden & B. Pincemin (eds.) } *JADT2008*, Presses Universitaires de Lyon, vol. 1, pp. 209-218.
- 13 Bolasco S. (2008a). *Corpora e liste di frequenza d'uso: criteri e tecniche per l'analisi automatica dei testi*. in Monica Barni, Donatella Troncarelli e Carla Bagna (eds.) "Lessico e apprendimenti. La dimensione lessicale nell'educazione linguistica", F. Angeli, Milano, pp. 113-142.
- 12 Bolasco S., Pavone P. (2007d). *Automatic dictionary and rule-based systems for extracting information from text*, in Classification and Data Analysis 2007. Book of short papers CLADAG2007. EUM - Edizioni Università di Macerata, pp. 255-258.
- 11 Bolasco S. (2007c). *Criteri e tecniche della statistica testuale per l'analisi automatica dei testi* in A. Grassi (ed.) *Demografia e Statistica: un ricordo di Enzo Lombardo tra scienza e cultura*, Dipartimento di Studi Geoeconomici Linguistici Statistici Storici per l'Analisi regionale, SAPIENZA Università di Roma, pp. 29-56.
- 10 Bolasco S., D'Avino E., Pavone P. (2007a) *Analisi dei diari giornalieri con strumenti di statistica testuale e text mining*, in Maria Clelia Romano (a cura di) "I tempi della vita quotidiana. Un approccio multidisciplinare all'analisi dell'uso del tempo." Roma, ISTAT, pp. 309-340.
- 09 La Noce M., Bolasco S., Allegra E., Ruocco V., Capo F.M. (2006b). Merger control in Italy 1995-2003: a statistical study of the enforcement practice by mining the text of Authority resolutions, in *International Journal of the Economics of Business* Vol. 13, 2, pp. 307-334.

- 08 Bolasco S., Galli de' Paratesi N., Giuliano L. (2006a). *Parole in libertà. Analisi statistica e linguistica dei discorsi di Berlusconi*, ManifestoLibri, Roma, pp. 142.
- 07 Bolasco S. (2005e). Statistica testuale e text mining: alcuni paradigmi applicativi, *Quaderni di Statistica*, Liguori Ed., 7, p. 17-53.
- 06 Baiocchi F., Bolasco S., Canzonetti A., Capo F. M. (2005d). Estrazione di informazione da testi per la classificazione automatica di una base documentale: la soluzione di Text Mining per l'Authority della Concorrenza, in S. Bolasco, et al. *Text mining: uno strumento strategico per imprese e istituzioni*, CISU, Roma, 2005, p. 45-54.
- 05 Bolasco S., Canzonetti A., Capo F. M. (2005d). *Text mining: uno strumento strategico per imprese e istituzioni*, CISU, Roma, pp. 202.
- 04 Bolasco S., Canzonetti A., Capo F. M., della Ratta-Rinaldi F., Singh B. K. (2005c) "Understanding Text Mining: a Pragmatic Approach" in S. Sirmakessis (ed.) *Knowledge Mining*, Series: Studies in Fuzziness and Soft Computing, Springer Verlag, Heidelberg, p. 31-51.
- 03 Bolasco S., Canzonetti A. (2005b). Some insights into the evolution of 1990s' standard Italian using Text Mining techniques and automatic categorisation, in M. Vichi, P. Monari, S. Mignani e A. Montanari (eds.) "New developments in classification and data analysis" Serie *Studies in Classification, Data Analysis, and Knowledge Organization*, Springer-Verlag, Berlin, p. 293-302.
- 02 Bolasco S. (2005a). La reperibilità statistica di tendenze diacroniche nell'uso delle parole, in I. Chiari e T. DeMauro (eds.) *Parole e Numeri. Analisi quantitativa dei fatti di lingua*, Aracne, Roma, p. 335-354.
- 01 Aureli E., Bolasco S. (a cura di) (2004). *Applicazioni di analisi statistica di dati testuali* Casa Editrice Università "La Sapienza", Roma, pp. 181.

**Quanto dichiarato nel presente curriculum vitae corrisponde al vero ai sensi degli artt. 46 e 47 del D.P.R. 445/2000.**

Roma, 25 febbraio 2015