



Roma 01/11/2011

Carissimi,

la presente è per annunciarvi la programmazione della seconda edizione della **Scuola internazionale di "Introduzione all'analisi automatica dei testi e ai metodi e modelli di text mining" (corso 2-ADT-TM)**, che si svolgerà a Roma nei mesi di febbraio (primi giorni) e aprile 2012. Visti gli ottimi risultati del primo corso, svoltosi lo scorso mese di gennaio, ci proponiamo di ripetere questa esperienza tenendo in considerazione la valutazione fatta dai partecipanti al termine della scuola. Quella verifica ci ha permesso di evidenziare due tipi di interlocutori: un primo gruppo di ricercatori più esperti, che ha operato una prima selezione circa i propri interessi, e un secondo gruppo più in fase di esplorazione, neofiti rispetto all'ambito specifico dell'analisi automatica dei testi o più in generale del trattamento statistico di dati non strutturati. Considerando le diverse aspettative di ciascun profilo, si è cercato di differenziare due percorsi tanto da privilegiare per uno piuttosto gli aspetti metodologici, per l'altro l'esplorazione delle aree e potenzialità applicative. Pertanto in questa nuova edizione, crediamo sia utile separare maggiormente i due aspetti per consentire a ciascuno di ottenere migliori risultati. In questa prospettiva, la scuola verrà organizzata secondo due moduli da svolgere in momenti successivi, intercalati fra loro da un periodo di due o tre mesi, per dare spazio di approfondimento e prime sperimentazioni a chi volesse seguire entrambi.

L'obiettivo del corso è di fornire strumenti di analisi statistica per la gestione di informazioni non strutturate, con un approccio multidisciplinare, con apporti provenienti da docenti linguisti, statistici, informatici, psicologi e sociologi.

I due moduli del corso saranno pertanto: 1) introduzione all'analisi automatica dei testi; 2) metodi, modelli e strumenti di text mining.

Nel primo modulo, si presentano le basi dell'analisi lessico-testuale del contenuto, che consentono di esplorare e interrogare raccolte di testi (basi documentali, rassegne stampa, interviste, domande aperte in indagini con questionario, forum, newsgroup, ...) . In particolare, si illustrano i processi di individuazione delle unità di analisi e di ponderazione delle parole in funzione degli obiettivi, nonché gli strumenti per la formulazione di modelli di estrazione di entità e di costruzione di risorse per il trattamento automatico. In questa fase, l'attenzione è rivolta in particolare alle applicazioni in campi differenti. Questi saranno individuati, una volta perfezionate le iscrizioni, tenendo conto degli interessi dei partecipanti. Il modulo prevede anche uno spazio laboratoriale per effettuare i primi passi operativi su software dedicati come Taltac, Lexico ed R.

Nel secondo modulo, si approfondiscono gli aspetti metodologici analizzando le tecniche sia classiche sia più recenti per la classificazione automatica, per la categorizzazione dei documenti e per l'estrazione di informazione, con attenzione alla visualizzazione del contenuto dei testi, alla validazione dei risultati e all'introduzione di informazioni esterne. Si illustrano, fra le altre, tecniche di *Sentiment Analysis* e di *Latent Semantic Analysis*. Gli esempi saranno tratti da applicazioni svolte con DTM-Vic, R e SAS.



La scuola è rivolta a dottorandi, ricercatori e studiosi di tecniche di analisi dei dati testuali o a quanti vogliono iniziare a occuparsi di questi temi.

Ogni modulo ha la durata di 21 ore distribuite su 3 giorni: 60% di lezione e 40% di pratica nell'uso di software e di illustrazione di casi di studio.

La scuola è organizzata dal dipartimento MEMOTEF (METodi e MOdelli per l'Economia, il Territorio e la Finanza) dell'Università "La Sapienza" di Roma e si svolgerà presso la facoltà di Economia, in via del Castro Laurenziano 9..

Fra i docenti figurano: Simona Balbi, Sergio Bolasco, Isabella Chiari, Francesca Dolcetti, Luca Giuliano, D. Fioredistella Iezzi, Ludovic Lebart, André Salem, Arjuna Tuzzi.

Il programma della prima edizione della scuola svoltasi nel gennaio scorso è disponibile sul sito web dei corsi del Dipartimento: <http://www.memotef.it/?pagina=corsi&tipo=3>.

Se pensate di voler partecipare al primo corso di questa seconda edizione, vi chiediamo di manifestarci entro il 20 novembre il vostro interesse inviandoci una mail all'indirizzo di sergio.bolasco@uniroma1.it oppure stella.iezzi@uniroma2.it (subject: **corso 2-ADT-TM**) indicando il vostro nome, cognome, mail e telefono.

Vi preghiamo inoltre di compilare il modulo di iscrizione disponibile sul sito, dove troverete ogni altra informazione utile. La quota di iscrizione vi sarà richiesta non appena il corso verrà confermato, avendo raggiunto il numero minimo di partecipanti.

La scheda dettagliata del programma sarà disponibile prossimamente sul sito web dei corsi del Dipartimento:

<http://www.memotef.it/?pagina=corsi&tipo=3>

Invitandovi a diffondere intorno a voi questo annuncio, prendiamo l'occasione per inviarvi i nostri più cordiali saluti

Stella Iezzi e Sergio Bolasco