

ASSOCIAZIONE PER  
L'INFORMATICA UMANISTICA  
E LA CULTURA DIGITALE



in collaborazione con

ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA



SCUOLA DI LETTERE E BENI CULTURALI  
DIPARTIMENTO DI FILOLOGIA CLASSICA E ITALIANISTICA  
DIPARTIMENTO DI INFORMATICA - SCIENZA E INGEGNERIA  
AIB - ASSOCIAZIONE ITALIANA BIBLIOTECHE

**AIUCD2014 - Terzo convegno annuale**  
*La metodologia della ricerca umanistica  
nell'ecosistema digitale*



**Book of abstracts**

Ed. by Federica Rossi and Francesca Tomasi

**AIUCD2014**

***La metodologia della ricerca umanistica nell'ecosistema digitale***

Abstracts 3rd annual conference, 18-19 settembre 2014

Biblioteca del Dipartimento di Filologia Classica e Italianistica (BFI)  
via Zamboni 32, Bologna

Edited by Federica Rossi and Francesca Tomasi

Associazione per l'Informatica Umanistica e la Cultura Digitale &  
University of Bologna - Scuola di Lettere e Beni Culturali, Dipartimento  
e Biblioteca di Filologia Classica e Italianistica, Dipartimento di  
Informatica - Scienza e Ingegneria.

Bologna, 2014

Ideazione e a realizzazione logo a cura di Marilena Daquino

# TABLE OF CONTENTS

## Il convegno

## L'Associazione

## Keynote address

Manfred Thaller. *Are the Humanities an endangered or a dominant species in the digital ecosystem?*

## Panel

### **Il ruolo delle Digital Humanities e le discipline umanistiche**

Raul Mordenti. *Procedura informatica e filologia*

Paola Moscati. *Verso il 2020: eredità e sfide dell'informatica archeologica*

Guido Abbattista. *Digital resources and historical research*

Claudio Gnoli. *Frammentazione delle informazioni e organizzazione della conoscenza*

## Papers

### **Metodologie a confronto**

Paola Italia e Fabio Vitali. *Varianti testuali e versioning fra bibliografia testuale e informatica*

Federico Condello e Mirko Degli Esposti. *La authorship attribution fra filologia e matematica*

Gianmaria Silvello e Karen Coyle. *Linked Open Data e biblioteche digitali*

## Papers

### **Arte, Storia e Letteratura**

Costanza Giannaccini, Susanne Müller e Francesca Di Donato. *Burckhardsource.org: a semantic digital library*

Maristella Agosti, Nicola Orio e Chiara Ponchia. *Guided Tours Across a Collection of Historical Digital Images*

Marilena Daquino. *Historical Context (HiCo): an ontology for describing context information of cultural heritage objects*

Silvia Stoyanova. *Remediating Leopardi's Zibaldone: harvesting and visualization of semantic networks in the scholarly archive*

Linda Spinazzè. *'Cursus in clausula': una risorsa online per lo studio ritmico della prosa*

## Papers

### **Linguistica, Filosofia e Scienze del documento**

Matteo Abrate, Angelo Mario Del Grosso, Emiliano Giovannetti, Angelica Lo Duca, Andrea Marchetti, Lorenzo Mancini, Irene Pedretti e Silvia Piccini. *Il Progetto Clavius on the Web: tecnologie linguistico-semantiche al servizio del patrimonio documentale e degli archivi storici*

Federico Boschetti, Riccardo Del Gratta, Claudia Marzi, Ouafae Nahli e Vito Pirrelli. *Modelli, metodi e strumenti per il trattamento automatico della lingua araba e per l'editing in ambienti collaborativi*

Simonetta Montemagni. *Modelli di variazione diatopica ed evidenza linguistica. Il contributo della dialettologia computazionale*

Pietro Giuffrida. *Dalla forma al contenuto. L'influenza della digitalizzazione dei testi filosofici sulla rappresentazione dei loro contenuti: il caso di Aristotele*

Cristina Marras. *Esplorando gli ambienti digitali per la ricerca in filosofia: sostenibilità e modelli di sviluppo*

## Tavola rotonda

### **I mestieri per le Digital Humanities**

Andrea Angiolini, Francesca Di Donato, Luca Rosati, Enrica Salvatori, Stefano Vitali

## Posters

Alessio Piccioli, Francesca Di Donato, Danilo Giacomi, Romeo Zitarosa e Chiara Aiola. *Linked Open Data per le Digital Humanities: l'esempio del Linked Open Data PORTAL*

Emanuele Di Buccio, Giorgio Maria Di Nunzio e Gianmaria Silvello. *Information Retrieval meets Geolinguistics: A Web Application for Measuring Distances Between Clitic Clusters*

Tiziana Mancinelli e Andrea Marchesini. *Un modello di ontologia per le strutture retorico-sintattiche di un testo lirico*

Marc Lasson e Corinne Manchio. *Misurare le parole: approccio digitale alla corrispondenza ufficiale di Machiavelli*

Antonella Brunelli e Roberto Brigati. *PhiloMed: un dialogo in rete tra Filosofia e Scienze Biomediche*

Marion Lamé, Chiara Mannari e Federico Ponchio. *Fonti in cerca di storici digitali: ridefinizione e mutamenti della fonte primaria*

Alessia Smaniotto. *OpenEdition: an international infrastructure for the digital humanities*

## **Il convegno**

Il terzo convegno annuale della Associazione per l'Informatica Umanistica e la Cultura Digitale - AIUCD 2014 - è dedicato al ruolo delle Digital Humanities nella ricerca delle discipline umanistiche. L'informatica ha sollecitato una ridefinizione dei processi teorici e metodologici nelle scienze umane, contribuendo anche ad una ri-concettualizzazione dei saperi delle discipline tradizionali. Arte, archeologia, filologia, filosofia, linguistica, scienze del libro e del documento, storia, ma anche discipline sociali e della comunicazione si avvalgono dei metodi della computer science per formalizzare i rispettivi ambiti di ricerca, per innovare pratiche e procedure. Ne consegue una profonda ridefinizione dei paradigmi disciplinari.

Alcuni concetti come Semantic Web, Linked Open Data, Digital libraries, archives and museums, information architecture, information visualization sono diventati elementi chiave anche della ricerca d'area umanistica.

Ecco allora che diventa urgente un confronto fra la ricerca delle discipline tradizionali e le Digital Humanities, per analizzare punti di convergenza e innovare metodi e strumenti in una prospettiva interdisciplinare e multidisciplinare.

### **L'organizzazione**

*Coordinatore generale*  
Dino Buzzetti

*Coordinatore del comitato di programma*  
Francesca Tomasi

*Comitato di programma*  
Maristella Agosti, Fabio Ciotti, Maurizio Lana, Federico Meschini, Nicola Palazzolo,  
Roberto Rosselli Del Turco, Anna Maria Tammaro

*Comitato scientifico locale*  
Luisa Avellini, Federica Rossi, Marialaura Vignocchi, Fabio Vitali

*Segreteria*  
Marilena Daquino, Federico Nanni

## L'Associazione

L'Associazione per l'Informatica Umanistica e la Cultura Digitale intende promuovere e diffondere la riflessione metodologica e teorica, la collaborazione scientifica e lo sviluppo di pratiche, risorse e strumenti condivisi nel campo dell'informatica umanistica, sia nell'uso delle applicazioni digitali in tutte le aree delle scienze umane, sia nel promuovere la riflessione sui fondamenti umanistici delle metodologie informatiche e nel campo delle culture di rete.

AIUCD è un'organizzazione associata alla European Association for Digital Humanities (EADH) ed è con ciò rappresentata nella Alliance of Digital Humanities Organizations (ADHO). Per raggiungere le proprie finalità l'Associazione persegue i seguenti obiettivi:

- ✓ promuove attività scientifiche e formative nel campo dell'informatica umanistica e delle applicazioni delle tecnologie digitali alle scienze umane;
- ✓ migliora la visibilità e la conoscenza di iniziative e risorse individuali, istituzionali e di altre associazioni, facilitando la concretizzazione di idee e progetti interdisciplinari;
- ✓ promuove l'integrazione delle competenze professionali umanistiche e informatiche di tutti gli interessati;
- ✓ stabilisce e consolida i legami con altre realtà europee ed internazionali che operano nel campo dell'informatica umanistica e delle applicazioni digitali alle scienze umane.

### Il direttivo

*Presidente*

Dino Buzzetti

*Vicepresidente*

Anna Maria Tammaro

*Tesoriere*

Nicola Palazzolo

*Consiglio Direttivo*

Maristella Agosti

Dino Buzzetti

Fabio Ciotti

Federico Meschini

Nicola Palazzolo

Maurizio Lana

Roberto Rosselli Del Turco

Anna Maria Tammaro

Francesca Tomasi

### Info e contatti

WEB - <http://www.umanisticadigitale.it>

FB - <http://www.facebook.com/groups/aiucd>

WIKI - <http://linclass.classics.unibo.it/udwiki/>

MAILING-LIST - [aiucd-l@humnet.unipi.it](mailto:aiucd-l@humnet.unipi.it)

BLOG - <http://infouma.hypotheses.org>

ASSOCIAZIONE PER  
L'INFORMATICA UMANISTICA  
E LA CULTURA DIGITALE



## Keynote address

### Manfred Thaller

#### *Are the Humanities an endangered or a dominant species in the digital ecosystem?*

The Humanities, like everybody else, live in a societal environment that is increasingly shaped by information technologies. The Digital Humanities are in this context occasionally seen as a trend which brings general salvation to the Humanities as a whole. This is certainly meaningless: The in some cases almost hysterical overtones, with which this argument is presented internationally is mainly influenced by the increasingly precarious position, which the Humanities have in the American – or possibly: Angloamerican – academic world. The arguments coming from this context are only partially applicable to Europe.

Nevertheless, as individual Humanities' disciplines over the centuries have been more or less important, the clearer their relevance for the large contemporary themes has been, the future of the Humanities will be shaped a bit by their relationship to the information society. Many Humanists believe, that their role within that society will be that of consumers of solutions developed by others. We will argue, that the Humanities can, if they are willing to face the challenge, become much more active players, influencing the future of this society built upon information.

### Profile

Professor in Computer Science for the Humanities at the Universität zu Köln.

Born 1950. Since 1970 studying history (originally history and Ancient Oriental Studies) at the University of Graz. There 1975 PhD in Modern History, with a thesis on "Studien zum Europäischen Amerikabild. Darstellung und Beurteilung der Politik und inneren Entwicklung der Vereinigten Staaten von Amerika in Deutschland, Großbritannien und Österreich zwischen 1840 und 1941 im Vergleich."

Afterwards postdoc fellowship in (empirical) sociology at the Institute for Advanced Studies in Vienna; here specialization in historical mobility studies. At the same time participation in research projects on the history of the family, study of the daily life of the Middle Ages and interlocking directorates of German and Austrian companies.

Since 1978 research fellow / senior research fellow at the Max-Planck-Institut für Geschichte, Göttingen. Responsible for the design and implementation of a general data base oriented programming system for history (CLIO / κλειω). At the same time research on a general methodology of historical computer science. Since 1995 also part-time professor at the University of Bergen, Norway. There responsible for teaching in «Historical computer Science».

September 1997 - February 2000 founding director, then permanent director of the «Humanities Information Technology Research Programme» and the attached research centre of the University of Bergen, Norway. Professor at the arts faculty there.

March 2000 onwards Prof. of «Historisch Kulturwissenschaftliche Informationsverarbeitung» at the University of Cologne, Germany.

## Panel

### Il ruolo delle Digital Humanities e le discipline umanistiche

#### **Raul Mordenti (Università di Roma "Tor Vergata")** ***Procedura informatica e filologia***

"L'edizione critica è un fatto complesso; essa è in realtà un processo consistente in diversi e sovrapposti atti interpretativi ed operazioni ermeneutiche, che vanno dalla prima lettura del testimone, attraverso la *recensio*, l'eventuale costruzione dello stemma, l'*emendatio*, la definizione del testo critico e del suo apparato, fino alla stampa del testo criticamente stabilito, e dunque ad una nuova lettura dato che, ormai lo sappiamo, l'organizzazione di ogni testo presuppone come proprio elemento costitutivo il lettore. Diciamo dunque che il lavoro dell'edizione critica descrive un *percorso* labirintico (ma, in linea di principio, razionalizzabile in forma di algoritmo) che congiunge due letture: una lettura L, quella del testo inedito da parte dell'editore (= start) e una lettura Lc, quella del testo criticamente stabilito da parte del destinatario dell'edizione (= stop)." (R. Mordenti, *Informatica a critica dei testi*, Bulzoni, Roma 2001, p. 54)

#### **Paola Moscati (ISMA - CNR, Roma)** ***Verso il 2020: eredità e sfide dell'informatica archeologica***

Se da un punto di vista teorico l'informatica archeologica del XXI secolo raccoglie ancora l'eredità delle esperienze maturate sin dagli anni '50 e '60 del Novecento, da un punto di vista metodologico le sfide più ardite per il futuro sono invece legate allo sviluppo esponenziale delle tecniche e degli strumenti di acquisizione, elaborazione e restituzione dei dati. La rassegna delle applicazioni informatiche si snoda seguendo le principali fasi dell'indagine archeologica, dalle ricerche sul campo, alle analisi di laboratorio, alla tutela e valorizzazione del patrimonio archeologico, tenendo conto anche degli aspetti legati alla promozione del turismo culturale e alle prassi operative integrate a esso collegate. In questa disamina, la rivista «Archeologia e Calcolatori», giunta al venticinquesimo anno di pubblicazione, costituisce un osservatorio privilegiato per analizzare e comprendere lo scenario attuale della ricerca e delineare le tendenze di sviluppo dell'informatica archeologica.

#### **Guido Abbattista (Università di Trieste)** ***Digital resources and historical research***

This contribution is devoted to explore in short the early 21<sup>st</sup> century scenario of online digital resources for historical studies as compared with the late 20<sup>th</sup> century situations and priorities. What promises have been fulfilled and what else can be expected in terms of real innovation in historical writing and publishing? Is it possible to assess the actual rate of change in academic historical research and teaching? What economic repercussions are visible in such connected fields as publishing and librarianship as far as historical knowledge production and dissemination are concerned? What is or could authorship become in present and future history writing? Which potentialities are at hand for a greater availability of historical information both for the general public and for what is commonly termed 'public history'? Finally a more general question will be addressed through a SWOT analysis of electronic publishing in connection with crucial issues such as scientific evaluation and career advancement.

#### **Claudio Gnoli (Università di Pavia; International Society for Knowledge Organization)** ***Frammentazione delle informazioni e organizzazione della conoscenza***

Insieme all'esplosione dell'informazione digitale, assistiamo a una sua crescente frammentazione. Le forme conchiusse nelle quali la conoscenza è stata fino ad ora veicolata -- monografie, trattati, riviste specializzate in un determinato dominio -- vengono affiancate o sostituite da unità sempre più brevi e temporanee -- moduli didattici, post, tweet, linked data -- che si disperdono attraverso la Rete in modi caotici. Se è vero che la modularizzazione stimola una più libera e creativa associazione (mash-up) delle informazioni e che la tecnologia ipertestuale facilita il loro collegamento e il loro recupero mediante i motori di ricerca, sul piano cognitivo si rischia tuttavia di smarrire la visione complessiva di ciascun tema e il reticolo di dipendenze concettuali che lo connette ad altri. Risulta dunque ancor più opportuno che nel passato il ricorso a sistemi di organizzazione della conoscenza (KOS), quali terminologie, tesauri, tassonomie, classificazioni, ontologie ecc., che permettano di tradurre l'interoperabilità tecnica dei dati in una vera interoperabilità concettuale. Diventa cioè tanto



necessaria quanto promettente la disciplina dell'organizzazione della conoscenza (knowledge organization, KO) che, a partire dalla storica esperienza di specializzazioni quali documentazione, biblioteconomia, linguistica e filosofia, negli ultimi decenni va generalizzando il proprio apparato di teorie, principi, metodi e strumenti; solo grazie a questi le scienze naturali ed umane potranno realizzare un'efficace sistematizzazione del patrimonio del sapere, mai assoluta né definitiva eppure sempre indispensabile al prosieguo della ricerca.

## Papers Metodologie a confronto

### **Paola Italia (Università di Roma "La Sapienza") e Fabio Vitali (Università di Bologna)** ***Varianti testuali e versioning fra bibliografia testuale e informatica***

Possibile che filologia e informatica possano collaborare non solo nella rappresentazione digitale delle variazioni dei testi nel tempo - le cosiddette Edizioni Critiche Digitali - ma anche nella caratterizzazione quantitativa e qualitativa delle loro varianti? Che conseguenze potrebbe avere un approccio critico alla bibliografia testuale, se integrata con le categorie del versioning e del change tracking? E nell'era degli scrittori nativi digitali, lo studio della varianti d'autore è destinato a scomparire, oppure l'informatica potrà fornire metodi e strumenti per indagare la genesi dei testi attraverso il text-logging?

A queste domande Paola Italia e Fabio Vitali provano a rispondere illustrando, attraverso un esempio di analisi variantistica qualitativa ed un prototipo costruito ad hoc per il celebre primo capitolo dei Promessi Sposi, i punti di contatto e le differenze metodologiche e culturali tra critica delle varianti e versioning dei testi. Per sviluppare una nuova metodologia di analisi, dai testi giuridici all'inventore dell'Azzeccagarbugli

### **Federico Condello (Università di Bologna) e Mirko Degli Esposti (Università di Bologna)** ***La authorship attribution fra filologia e matematica***

Nel campo della *authorship attribution*, si assiste da tempo a una proliferazione di approcci quantitativi fondati sia su metodi propri della linguistica computazionale e del *machine learning*, sia su metodi propri della fisica-matematica, che scelgono quale base d'analisi non le 'parole', ma unità più astratte come gli *n-grams*. Concetti matematici tipici della 'teoria dei sistemi dinamici' e della 'teoria dell'informazione' (e.g. la nozione di entropia relativa e gli aspetti matematici relativi agli algoritmi di compressione dati) offrono strumenti computazionali efficaci per l'attribuzione di un testo acefalo entro un numero finito di autori possibili. Molto più difficile, ma perciò più stimolante, l'applicazione di tali strumenti a un testo parzialmente o integralmente falso, perché in tal caso la 'similarità' stilistica è un argomento del tutto controvertibile. A partire da un *case-study* illustre, da poco tornato agli onori della cronaca (il cosiddetto *Diario postumo* attribuito incautamente a Eugenio Montale), Federico Condello (filologo classico) e Mirko Degli Esposti (matematico) spiegano la costruzione e gli esiti degli esperimenti condotti (con la collaborazione di Dario Benedetto [Roma]) a ricalzo delle tecniche filologiche più tradizionali; esperimenti che acquisiscono rilevanza metodologica anche al di là dei loro concreti e pur interessanti risultati.

### **Gianmaria Silvello (Università di Padova) e Karen Coyle (già University of California)** ***Linked Open Data e biblioteche digitali***

La conversazione si concentra su convergenze e divergenze tra cataloghi di biblioteca come sistemi chiusi e sistemi informativi basati su Internet ed il Web. La tecnologia dei LOD (Linked Open Data) sembra colmare il gap tra il paradigma tecnico e la possibilità di facilitare la creazione di conoscenza ma ancora esistono divergenze. Le domande che evidenziano il diverso approccio metodologico della comunità dei bibliotecari e quella degli informatici sono: strings vs things? LOD: creati e cercati manualmente vs LOD creati e cercati con tecniche automatiche? citazione e provenienza dei dati vs. mappatura da vari formati e strutture? la comunicazione tra le due comunità finora è stata assente o è stata superficiale e dovrà essere continuata.

## Papers

### Arte, Storia e Letteratura

**Costanza Giannaccini (SNS - Pisa), Susanne Müller (SNS - Pisa) e Francesca Di Donato (SNS - Pisa e Net7 - Pisa)**

#### ***Burckhardsource.org: a semantic digital library***

Burckhardsource.org is a semantic digital library created within the project "The European correspondence to Jacob Burckhardt". It concentrates on a time span ranging from 1842 to 1897 and witnesses a period of significant cultural transformations. After a brief digression on the technologies used in the diplomatic edition, from manuscripts transcription to the transformation of ODT encoded transcriptions in XML files, the focus will move towards specific technical tools created to add information to texts thanks to semantic annotation. In particular the use of Pundit, an innovative tool to annotate web contents based on Linked Data technologies, and the Semantic Web basket manager Korbo. On the wake of semantic annotation, a final consideration will touch upon the enrichment of the web of data through the addition of new entries in Freebase.

**Maristella Agosti (Università di Padova), Nicola Orio (Università di Padova) e Chiara Ponchia (Università di Padova)**

#### ***Guided Tours Across a Collection of Historical Digital Images***

One of the challenges in making available a digital collection of cultural content is raising and maintaining the interest of non-specialized users. To this end, we propose to enrich the functions of a digital archive with short lectures, called narratives, as a personalization method to support and guide users inside a collection of cultural material. Narratives aim at highlighting the richness of the available digital content, which may be difficult to retrieve for non-experts; yet users are still free to interact with the collection according to their own interests and using all the system tools. We tested the approach with a task-oriented evaluation. A group of users followed a narrative before interacting with the content, while a control group accessed the content directly. Results showed that, since narratives included an introduction on the usage of all the available tools, the user group that followed the guided tour was facilitated in learning how to interact with the system and access to relevant content.

**Marilena Daquino (Università di Bologna)**

#### ***Historical Context (HiCo): an ontology for describing context information of cultural heritage objects***

The general approach in cultural heritage objects description reflects a common need to adhere to interoperability standards, preserving however the richness of data representation. Therefore, the common interest is to converge on a data-centric approach in order to exchange among communities useful information to clearly identify objects and their context, where context is a semantic deep keyword for referring to any information describing not only the object, but also the network of relations in which it's somehow involved.

This compels communities an effort to define a formalized, shared and exchangeable model. Semantic Web technologies are conceived in order to achieve this goal, seeking for (desirable) shared data models, by guaranteeing moreover the reuse of referenced models. For these reasons HiCO, an OWL 2 DL ontology for describing the context of cultural heritage objects, has been formalized. A special attention has been given to the concept of cultural heritage object itself: any information extracted by an agent from the "content" of the object, or from any other source dealing with the object, in an historical outlook, are entities bounded to the object through interpretation acts. So, a particular focus has been directed to event-driven and interpretation-driven approaches, in order to formalize such complex information in a Linked Data perspective.

**Silvia Stoyanova (Trier Center for Digital Humanities - Germany)**

#### ***Remediating Leopardi's Zibaldone: harvesting and visualization of semantic networks in the scholarly archive***

The project of remediating Giacomo Leopardi's Zibaldone exploits its hypertextual dimension of cross-references between related passages and of their allocation to thematic indexes, which Leopardi wrote for the purpose of mediating his intellectual diary into a more expository mode of discourse. The

project's objectives are to actualize the author's design for harvesting the Zibaldone's intra-textual semantic networks and to build a hypertext research platform with editorial apparatus, which would enable users to mine the text's structural complexity more comprehensively than is possible with a print edition or with its digital transcription. The XML encoding in TEI5 allows to access the Zibaldone modularly, in custom-selected layers of its encoded elements and explore their interrelations through statistical charts, histograms, visualizations. The platform aspires to become a collaborative space, where users could add their own annotations to the text and contribute to the site's apparatus, such as adding the linked contents of its bibliographic references, a database of its critical bibliography, etc. The project is a collaboration between Princeton University and the Trier Center for Digital Humanities.

**Linda Spinazzè (Università di Udine)**

***'Cursus in clausula': una risorsa online per lo studio ritmico della prosa***

Lo studio del cursus in un testo latino può essere uno strumento importante per la critica testuale e può fornire addirittura un sostegno in questioni attributive o riguardanti l'epoca di composizione. Tuttavia la percezione di questo elemento stilistico 'soprasegmentale' (Chiesa 2012) soprattutto quando si tratta di cursus prosodico (classico o mixtus) è particolarmente ostico per noi moderni che abbiamo perso la facoltà di distinguere a orecchio sillabe brevi e lunghe e non abbiamo idea di come fosse davvero declamato il latino letterario in un'età dove l'oralità e la lettura ad alta voce erano il modo più comune di fruizione del testo. Studiare questo aspetto retorico è attività che richiede dapprima una scansione metrica e ritmica di ogni clausola (se non anche di cola all'interno del periodo) e poi l'applicazione di metodi statistici di comparazione per riuscire a valutare oggettivamente l'intenzionalità o meno dell'autore nel manipolare il ritmo. La risorsa online Cursus in clausula mira dunque ad aiutare lo specialista sia nella prima fase 'meccanica' di scansione metrica e ritmica delle clausole, che nella seconda fase di riorganizzazione statistica dei risultati e calcolo del  $\chi^2$  secondo il metodo della comparazione interna inaugurato da Tore Janson (1975) per il cursus ritmico medievale, e poi variamente teorizzato e messo in pratica sulla letteratura di transizione (cursus mixtus) da Orlandi e la sua scuola. Trattando semplici file di testo caricati dall'utente e applicando le regole delle fonosintassi latina ad ogni gruppo di parole posto in prossimità di un segno di punteggiatura forte, il programma distingue sillabe brevi e sillabe lunghe e rileva la cadenza dell'accento proponendo così in output sia una scansione prosodica (tipo: ++-|+\*) sia una scansione ritmica (tipo: 3p p) di tutte le clausole.

## Papers

### Linguistica, Filosofia e Scienze del documento

**Matteo Abrate (IIT-CNR - Pisa), Angelo Mario Del Grosso (ILC-CNR - Pisa), Emiliano Giovannetti (ILC-CNR - Pisa), Angelica Lo Duca (IIT-CNR - Pisa), Andrea Marchetti (IIT-CNR - Pisa), Lorenzo Mancini (ILC-CNR - Pisa), Irene Pedretti (APUG - Roma) e Silvia Piccini (ILC-CNR - Pisa)**

***Il Progetto Clavius on the Web: tecnologie linguistico-semantico al servizio del patrimonio documentale e degli archivi storici***

L'obiettivo del progetto Clavius on the Web è quello di valorizzare una parte dei manoscritti conservati dall'Archivio storico della Pontificia Università Gregoriana (APUG). Il progetto prende in esame alcuni documenti (un corpus di oltre 300 lettere e 7 manoscritti in parte inediti) relativi a Christophorus Clavius (1538-1612), matematico e astronomo gesuita. Il carteggio è stato digitalizzato e trascritto ed alcune lettere in latino sono state tradotte e analizzate dal punto di vista linguistico, lessicale e semantico. In particolare, attraverso l'analisi linguistica, ad ogni unità ortografica (token) che compone il testo sono state associate una lettura morfologica ed un lemma. Tale fase è stata realizzata attraverso l'utilizzo sia di strumenti e modelli stocastici per l'attribuzione automatica della lettura morfologica (Hidden Markov Model) sia di autorevoli repertori linguistici (formario-lemmario e treebank del Perseus Project) per l'attribuzione del lemma. Nella fase di analisi lessicale, quindi, sono stati individuati nel carteggio tutti i termini chiave appartenenti al dominio astronomico e matematico ed il loro contenuto semantico è stato rappresentato e formalizzato in un lessico-thesaurus elettronico. In quest'ultimo sono stati formalizzati anche tutti quei concetti discussi e definiti da Clavius nella sua

Opera Mathematica (1612). Infine, per quanto concerne l'analisi semantica, sono state annotate nel testo le entità rilevanti come i personaggi storici, i luoghi, gli eventi ed i riferimenti bibliografici, per fornire una ulteriore chiave di accesso al contenuto dei manoscritti. L'integrazione fra questi tre livelli di analisi è stata realizzata attraverso l'uso di CTS URN. La terminologia e le entità di dominio individuate nel testo sono state inoltre strutturate in un lessico e una ontologia, esportate come Linked Data e collegate a risorse già disponibili sulla Rete, come ad esempio VIAF, DBpedia, Wikipedia e GeoNames. Attualmente, il catalogo delle risorse contiene personaggi, luoghi e lettere associati al dominio di Clavius. La loro visualizzazione sul Web è implementata mediante due differenti tecniche di Data Visualization: la prima si basa sulla corrispondenza tra testo, traduzione e immagine mentre la seconda ha lo scopo di mostrare i contenuti delle analisi, con particolare enfasi alla navigazione delle risorse lessicali e ontologiche prodotte durante le varie fasi.

**Federico Boschetti (ILC-CNR - Pisa), Riccardo Del Gratta (ILC-CNR - Pisa), Claudia Marzi (ILC-CNR - Pisa), Ouafae Nahli (ILC-CNR - Pisa) e Vito Pirrelli (ILC-CNR - Pisa)**

***Modelli, metodi e strumenti per il trattamento automatico della lingua araba e per l'editing in ambienti collaborativi***

La linguistica computazionale ha portato negli ultimi vent'anni a un profondo mutamento nello studio delle lingue e delle loro testimonianze scritte, spostando l'accento della ricerca da aspetti linguistico-formali all'uso linguistico in contesti comunicativi reali. Il presente contributo illustra l'impatto di questo cambio di prospettiva sullo studio della lingua araba, attraverso una rassegna di alcune attività di ricerca in corso presso l'Istituto di Linguistica Computazionale del CNR di Pisa:

- I. acquisizione dei testi arabi tramite Optical Character Recognition (OCR) e sviluppo di strumenti per la correzione manuale del testo in ambienti collaborativi;
- II. sviluppo di algoritmi e strumenti per l'analisi morfologica della lingua araba;
- III. analisi delle dinamiche di acquisizione del lessico arabo mediante architetture bio-computazionali;
- IV. sviluppo della WordNet dell'Arabo collegata a Princeton WordNet, ItalWordNet, LatinWordNet e alla nascente AncientGreek WordNet.

Queste attività sono rivolte sia all'analisi delle caratteristiche linguistiche dell'arabo che allo studio della produzione letteraria araba e dei suoi rapporti storico-culturali con altre lingue. In particolare, il contributo intende illustrare la fertilità di un approccio metodologico che metta in relazione le dinamiche di acquisizione del lessico arabo, con la messa a punto di procedure di analisi ed edizione critica del testo e con i principi di organizzazione ontologica di una lingua ad alta produttività derivazionale.

**Simonetta Montemagni (ILC-CNR - Pisa)**

***Modelli di variazione diatopica ed evidenza linguistica. Il contributo della dialettologia computazionale***

Il presente contributo intende indagare se e in che misura i risultati di studi dialettologico-computazionali possano contribuire a fare avanzare filoni tradizionali di ricerca in dialettologia. Focalizzandoci sul tipo di connessione tra il modello di variazione diatopica ipotizzato e l'evidenza linguistica offerta dal corpus di materiali dialettali, verrà indagato se sia realmente possibile formulare modelli della variazione diatopica che riflettano tendenze generali e al contempo mantengano un legame diretto e deterministico con i dati raccolti sul campo. Questi hanno rappresentato finora requisiti inconciliabili: se da un lato modelli della variazione diatopica costruiti induttivamente si sono limitati a pochi e accuratamente selezionati tratti linguistici, dall'altro modelli generali della variazione dialettale costruiti deduttivamente non risultano essere completamente verificabili. Oggi, il ricorso a metodi e tecniche computazionali di analisi del dato dialettale può aiutare a superare i limiti intrinseci degli approcci delineati sopra e a combinarli in maniera produttiva e innovativa. Ripercorrendo i passi intrapresi per l'analisi automatica del corpus dei materiali dialettali dell'*Atlante Lessicale Toscano* (ALT), il contributo mostrerà come metodi e tecniche di analisi computazionale del dato dialettale sono oggi in grado di soddisfare questo duplice requisito, aprendo nuovi e inesplorati orizzonti sul versante della ricerca dialettale.

**Pietro Giuffrida (Università di Palermo)**

***Dalla forma al contenuto. L'influenza della digitalizzazione dei testi filosofici sulla rappresentazione dei loro contenuti: il caso di Aristotele***

L'applicazione dell'informatica umanistica allo studio del corpus aristotelico può permettere di aggiornare la rappresentazione complessiva delle intenzioni teoriche perseguite dallo Stagirita. Uno degli indizi relativi alla configurazione complessiva del corpus aristotelico, e quindi al problema della genesi delle singole opere e dei loro rapporti di dipendenza teorica, è costituito dai rimandi interni che, all'interno dei singoli trattati, collegano reciprocamente gran parte delle opere sopravvissute. Il progetto che si intende presentare ha come scopo 1. la classificazione, all'interno di un database relazionale appositamente progettato, della totalità dei riferimenti interni al corpus aristotelico; 2. la costruzione di un'interfaccia di consultazione dei dati raccolti; 3. l'elaborazione di una descrizione del corpus aristotelico, ovvero di una rappresentazione complessiva dei temi, dei nuclei teorici e dei metodi, a partire dal tessuto dei rimandi interni, cui si attribuisce funzione analoga a quella degli hyperlink.

Per esemplificare il metodo e gli obiettivi del progetto, saranno presentati i primi risultati di un progetto pilota sviluppato sul testo del *De motu animalium*, che in proporzione alla sua esigua estensione contiene un numero eccezionalmente grande di rinvii ad altri trattati.

**Cristina Marras (ILIESI-CNR - Roma)**

***Esplorando gli ambienti digitali per la ricerca in filosofia: sostenibilità e modelli di sviluppo***

L'Umanistica Digitale, per definizione, è una disciplina 'ibrida', nata in un certo qual modo dall'incontro e dal non certo facile matrimonio fra tradizioni e pratiche disciplinari differenti.

La presentazione si concentra sul lavoro condotto per la pubblicazione digitale di un manoscritto del filosofo tedesco Gottfried Wilhelm Leibniz (1676-1716), un caso studio al fine di mettere in evidenza le componenti che concorrono all'organizzazione del lavoro di ricerca in un ambiente digitale e alle possibilità e ai limiti offerti dagli ambienti integrati per la ricerca collaborativa e la pubblicazione e alle loro connesse molteplici funzionalità.

Vengono problematizzati i modi attraverso i quali tecnologie e linguaggi influenzano il lavoro sul testo e le conseguenze di questo per le pratiche di ricerca e per le relazioni tra ricercatori nella cornice di un processo di mutua ridefinizione dei paradigmi disciplinari.

Si descrive e si esplora dunque un 'micro' ecosistema digitale e, a partire da un esempio concreto di lavoro che coniuga sia teoria e pratica, sia tecnologie e strumenti tradizionali di analisi del testo, si riflette sulle metodologie e soprattutto sulla conoscenza multidisciplinare che si produce e i linguaggi metaforici che più adeguatamente possono coglierla, descriverla e rappresentarla.

## **Tavola rotonda I mestieri per le Digital Humanities**

**Andrea Angiolini (Il Mulino, Bologna), Elogio dell'ornitorinco**

Dal punto di vista di un editore, le competenze collegate alle DH devono trovare un punto di equilibrio tra l'approccio culturale e quello tecnologico, distinguendosi al contempo dalla semplice alfabetizzazione digitale. Infatti, i diversi profili dovranno poter contare su solide basi culturali alle quali aggiungere conoscenze tecnologiche, dimestichezza con linguaggi e formati, pratica delle dinamiche sociali e commerciali in rete e basi legali. Che, in tutto questo, possa aiutare l'ornitorinco?

**Francesca Di Donato (SNS Pisa e Net7 srl - Pisa)**

L'intervento sarà orientato a illustrare come Net7 (<http://www.netseven.it>), un'azienda software specializzata in applicazioni web per le Humanities e coinvolta in numerosi progetti europei nell'ambito del Cultural heritage e delle Digital Humanities, integra le competenze umanistiche e informatiche nella realizzazione dei suoi prodotti. A partire da alcuni progetti ed esperienze concrete, si mostrerà come le competenze che si acquisiscono nei corsi di studi che fondono la formazione nel campo delle tecnologie digitali e in ambito umanistico, possono trovare concreta applicazione in nuovi mestieri.

**Luca Rosati (Information Architecture - Perugia)**

“Information is not knowledge” canta Frank Zappa in *Packard Goose*, facendo il verso a un famoso verso di Eliot. La società contemporanea è caratterizzata da un sovraccarico informativo senza precedenti: di informazione ne abbiamo fin troppa, ciò che ci serve sono strumenti per renderla fruibile, per renderla cioè capace di diventare conoscenza. Questo è possibile fornendo un contesto all’informazione stessa, collocandola all’interno di un sistema di relazioni in grado di generare senso. È quanto fa l’architettura dell’informazione. Non una nuova disciplina, ma un sapere di confine che nasce all’intersezione di saperi vecchi e nuovi. L’architettura dell’informazione permette di migliorare e arricchire la nostra interazione con l’informazione in tutti gli spazi informativi condivisi: digitali (siti web, intranet, banche dati), fisici (come punti vendita, musei, ospedali), procedurali (flussi informativi nei processi aziendali).

### **Enrica Salvatori (Università di Pisa)**

Il corso di laurea di Informatica Umanistica a Pisa vanta ormai 12 anni di vita. L’intervento cercherà di evidenziare successi e problematiche sul piano dell’impiego lavorativo dei dottori di Informatica Umanistica (triennale e magistrale) sia guardando ad alcuni dati statistici sia focalizzando l’attenzione su alcune storie particolari. Si cercherà di capire in quali ambiti i laureati di Informatica Umanistica hanno trovato maggiore impiego, quali i motivi di carriera e le resistenze, quali prospettive potrebbero delinearsi.

Rimane poi sul fondo la questione dell’identità, l’essere informatico umanista nel mondo del lavoro, di cui si cercherà di delineare i margini.

### **Stefano Vitali (Soprintendenza archivistica per l’Emilia Romagna - Bologna)**

L’avvento del digitale ha comportato una sorta di processo di espropriazione di saperi e competenze tecniche tradizionalmente tipiche del mestiere dell’archivista. Non a caso si sente spesso dire che oggi gli strumenti di ricerca sono un prodotto più dei software di inventariazione che delle scelte interpretative e delle strategie comunicative dell’archivista, così come l’organizzazione degli archivi correnti è funzione dei sistemi di protocollo e di gestione del flusso documentario piuttosto che delle autonome valutazioni dell’archivista, in relazione ai bisogni di autodocumentazione dei soggetti che sedimentano gli archivi. Il risultato di questo processo non è soltanto un isterilimento dei connotati umanistici e storico-critici del mestiere dell’archivista a beneficio delle sue componenti puramente tecnicistiche, quanto anche un tendenziale appiattimento della qualità dei prodotti descrittivi e difficoltà crescenti nella organizzazione degli archivi correnti. Questo processo può essere contrastato anche grazie ad una maggiore presenza all’interno dei processi formativi della professione di conoscenze teoriche e di abilità informatiche che permettano all’archivista di porsi in maniera critica di fronte ai prodotti che si trova ad utilizzare – quasi sempre non per sua scelta diretta – e di svincolarsi, laddove se ne presenti la necessità e l’opportunità, dal ricorso a software imposti da un mercato non sempre attento alle reali necessità di chi si trova a farne uso.

## **Posters**

### **Alessio Piccioli (Net7 srl - Pisa), Francesca Di Donato (SNS Pisa e Net7 srl - Pisa), Danilo Giacomi (Net7 srl - Pisa), Romeo Zitarosa (Net7 srl - Pisa) e Chiara Aiola (Net7 srl - Pisa)** ***Linked Open Data per le Digital Humanities: l’esempio del Linked Open Data PORTAL***

Il Linked Open Data PORTAL (o LODPORTAL) è stato sviluppato all’interno del progetto europeo europeo (CIP-ICT-PSP) Agorà – Scholarly Open Access Research in European Philosophy. Pur essendo ancora un prototipo, l’applicazione, rilasciata con licenze open source, presenta un uso innovativo delle tecnologie Linked Data applicate all’ambito delle Digital Humanities. Associata all’impiego di tecnologie del Semantic Web, la combinazione di risorse provenienti da diversi content provider offre al ricercatore strumenti innovativi per la creazione, il riuso, la fruizione e la visualizzazione dei contenuti. In particolare, il LODPORTAL raccoglie e indicizza risorse provenienti da dieci biblioteche digitali, create dai gruppi di ricerca afferenti al progetto Agorà e dalla cosiddetta Linked Open Data Cloud (dbpedia e freebase) e include un modulo, basato su Pundit, per l’estrazione automatica di entità e l’arricchimento semantico manuale. Grazie all’indicizzazione semantica, l’interfaccia di navigazione offre alcune visualizzazioni avanzate, tra cui la timeline, il grafo delle persone e una mappa dei luoghi citati e associati alle risorse delle biblioteche digitali.

**Emanuele Di Buccio (Università di Padova), Giorgio Maria Di Nunzio (Università di Padova) e Gianmaria Silvello (Università di Padova)**  
***Information Retrieval meets Geolinguistics: A Web Application for Measuring Distances Between Clitic Clusters***

Syntactic variation studies the ways in which linguistic elements, i.e. words and clitics, are put together to form constituents, that are phrases or clauses. In the field of linguistics, syntactic comparison across closely-related languages has proven to be extremely useful for several tasks, such as: finding relations between cross-linguistic syntactic differences that might otherwise appear unrelated; analysing the linguistic structures in the task of historical reconstruction. In this paper, we present the ongoing research on the use models from the field of Information Retrieval, in particular the Vector Space Model, to represent dialects in a multidimensional space and measure both the syntactic and the geographical distance between dialects by means of the definition of geometric distance in a vector space.

**Tiziana Mancinelli (University of Reading - UK) e Andrea Marchesini (Firefox - UK)**  
***Un modello di ontologia per le strutture retorico-sintattiche di un testo lirico***

Il progetto propone un modello di ontologia per la rappresentazione delle figure retoriche e di aspetti sintattici in un testo poetico (ampliabile anche in altri documenti) e un *tool* che possa supportare il processo di codifica e di annotazione facilitando il lavoro dell'editore. Lo scopo principale è quello di fornire uno strumento di mappatura e di descrizione formale dei testi che possa essere applicato e collegato a documenti TEI, flessibile per le annotazioni multiple e per coniugare gli aspetti peculiari di ogni testo con la rappresentazione stilistica di aspetti retorico-sintattici. L'obiettivo è quello di suggerire un'ontologia basata sulle specifiche del W3C e del Semantic Web utilizzabile per la mappatura di figure retoriche e aspetti sintattici capace di relazionarsi con ogni ontologia esistente, dal DublinCore alla TEI/ CIDOC-CRM, e in grado di essere espansa e usata in ogni forma di documento secondo quanto prevedono gli standard del Semantic Web, RDF, RDFs e OWL.

**Marc Lasson (INRIA - Paris) e Corinne Manchio (Université Paris 8)**  
***Misurare le parole: approccio digitale alla corrispondenza ufficiale di Machiavelli***

Questo intervento si basa sulla creazione di un software adatto all'analisi della corrispondenza ufficiale di Machiavelli. Le Legazioni e Commissarie sono il punto di partenza di un lavoro di dottorato sulla lingua del Segretario fiorentino tra il 1498 e il 1512. Gli strumenti informatici ci hanno permesso non solo di ordinare ed esplorare un materiale molto ricco (2257 lettere) ma anche di sfruttarlo e di scoprire proprietà ed usi della lingua specifici. Nel nostro progetto, lo strumento informatico non viene concepito come un software da rilasciare ma come uno strumento di esplorazione dei testi sufficientemente elastico per adattarsi rapidamente sia ai bisogni dei ricercatori sia alla verifica di nuove ipotesi. Cercheremo di esporre qualche risultato (sotto varie forme, diagrammi, strutture arboreescenti) a partire dagli usi della parola guerra.

**Antonella Brunelli (Università di Bologna) e Roberto Brigati (Università di Bologna)**  
***PhiloMed: un dialogo in rete tra Filosofia e Scienze Biomediche***

Concepito e alimentato da un gruppo di lavoro in cui si intrecciano numerose competenze, scientifiche e tecniche, il Portale PhiloMed si propone come uno strumento di orientamento nel dialogo tra Filosofia e Scienze mediche. Andando oltre il semplice styling grafico accattivante il portale rinnova gli schemi delle biblioteche digitali classiche mediando tra il linguaggio comunicativo del web narrativo (emozionale) e quello del web di reference (di consultazione). Oggetto di confine tra cataloghi tradizionali e Virtual Reference Desk tematico, Philomed struttura contenuti eterogenei (eventi, attività, documenti) e documentazione multimediale (libri e articoli, foto, video, etc.) intorno a "focus tematici" indagati in modo multidisciplinare e rappresentati in mappe concettuali in forma di grafo. Tutti i contenuti sono indicizzati attraverso metadati strutturati ovvero in ontologie. Prossimi obiettivi: lavoro sull'indicizzazione per la creazione di un dizionario interno e mappe semantiche derivati da thesauri e dizionari disciplinari controllati, (seguendo le indicazioni degli standard Dublin Core), abilitazione al commenting e ai social, creazione di gruppi di utenti con privilegi riservati per l'elaborazione condivisa di materiali di lavoro, segnalazione agli utenti delle modalità e policies di

accesso ed utilizzo dei documenti attraverso la creazione di pulsanti distintivi assegnati a ciascun record.

**Marion Lamé (ILC-CNR - Pisa), Chiara Mannari e Federico Ponchio (ISTI-CNR - Pisa)**  
***Fonti in cerca di storici digitali: ridefinizione e mutamenti della fonte primaria***

Questo contributo si propone di presentare il quadro complessivo, sia teorico e metodologico, che pragmatico, di una ricerca in informatica umanistica messa al servizio della storia e dell'epigrafia e nutrita dalle metodologie tradizionali di queste due discipline. Questa ricerca si applica al caso di studio del progetto TSS (Tesserarum Sisciae Sylloge), edizione scientifica digitale di iscrizioni concepita come strumento complementare al discorso storico e letterario dello studioso, mettendo al centro del processo di digitalizzazione la fonte e il suo storico. Si presenteranno le scelte di navigazione e le conseguenze implementative dell'oggetto epigrafico inteso come fonte storica e gli aspetti relativi alla progettazione assieme allo sviluppo dell'interfaccia di accesso ai dati strutturati nel database, delle applicazioni di visualizzazione per mettere a disposizione un Dispositivo Autoptico Digitale (TSS viewer e RTI viewer) e del sito che ospita il progetto.

**Alessia Smaniotto (Cléo - OpenEdition)**

***OpenEdition: an international infrastructure for the digital humanities***

OpenEdition is a comprehensive infrastructure for online books, journals, academic programs and scholarly blogs in the humanities and social sciences. It's developed by the Center for Open Electronic Publishing (Cleo), a non-profit public initiative promoting Open Access academic publishing, with the support of major French research institutions.

OpenEdition develops R&D programs in Digital Humanities based on text and data mining technologies.

OpenEdition develops a new model for communication and publication in humanities and social sciences based on the open web and social media. It aims also to create a disruptive economic model for Open Access based on licensing premium services to libraries and individual readers. OpenEdition brings together four complementary platforms dedicated to journals (Revue.org), books series (OpenEdition Books), research blogs (Hypotheses) and academic announcements (Calenda), each one implemented with a wide range of services dedicated to researchers, publishers and libraries.

Involved in the European DH Community (with contribution to Dariah, Centernet Europe and the Humanistica association), OpenEdition also contributes to the dissemination of the DH principles and uses by organizing seminars and editing a books collection on these topics. OpenEdition was deeply instrumental in the conception and dissemination of the DH Manifesto.