



Università  
di Catania

ASSOCIAZIONE per  
l'INFORMATICA UMANISTICA  
e la CULTURA DIGITALE

Consiglio Nazionale  
delle Ricerche

# ME.TE. DIGITALI

## MEDITERRANEO IN RETE TRA TESTI E CONTESTI

ATTI DEL XIII CONVEGNO ANNUALE  
AIUCD 2024



28 - 30 MAGGIO  
MONASTERO DEI BENEDETTINI  
P.ZZA DANTE, 32 CATANIA

ISBN 978-88-942535-8-0



Copyright ©2024 AIUCD  
Associazione per l'Informatica Umanistica e la Cultura Digitale



Il presente volume e tutti i contributi sono rilasciati sotto licenza Creative Commons Attribution Share-Alike 4.0 International license (CC-BY-SA 4.0). Ogni altro diritto rimane in capo ai singoli autori.  
*This volume and all contributions are released under the Creative Commons Attribution Share-Alike 4.0 International license (CC-BY-SA 4.0). All other rights retained by the legal owners.*

A cura di: Di Silvestro Antonio; Spampinato Daria (2024). Me.Te. Digitali. Mediterraneo in rete tra testi e contesti, Proceedings del XIII Convegno Annuale AIUCD, Catania 28-30 maggio 2024, Università di Catania.

Editing: Denise Bruno; Christian D'Agata; Laura Mazzagufò; Francesca Prado; Eliana Vitale; Alessandro Zammataro.

Ultimo accesso agli URL in data 15 maggio 2024.

Si prega di notificare all'editore ogni omissione o errore si riscontri: [aiucd.segreteria \[at\] aiucd.org](mailto:aiucd.segreteria@aiucd.org)  
*Please notify the publisher of any omissions or errors found: [aiucd.segreteria \[at\] aiucd.org](mailto:aiucd.segreteria@aiucd.org)*

Il programma della conferenza AIUCD 2024 è disponibile online <https://aiucd2024.unict.it/programma/>  
*The AIUCD 2024 Conference Program is available online <https://aiucd2024.unict.it/programma/>*

I contributi pubblicati nel presente volume hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima mediante *double-blind peer review* sotto la responsabilità del Comitato di Programma di AIUCD 2024.

*All the papers published in this volume have received favourable reviews by experts in the field of DH, through an anonymous double-blind peer review process under the responsibility of the AIUCD 2024 Program Committee.*

## **Chair**

Antonio Di Silvestro (Università di Catania)

Daria Spampinato (CNR Istituto di Scienze e Tecnologie della Cognizione)

## **Comitato di programma / Program committee**

Emmanuela Carbé (Università di Siena)

Massimo Cultraro (CNR Istituto di Scienze del Patrimonio Culturale)

Christian D'Agata (Università di Catania)

Antonio Di Silvestro (Università di Catania)

Greta Franzini (Eurac Research)

Maurizio Lana (Università del Piemonte Orientale)

Cristina Marras (CNR Istituto del Lessico intellettuale europeo e Storia delle Idee)

Marco Mazzone (Università di Catania)

Ouafae Nahli (CNR Istituto di Linguistica Computazionale "Antonio Zampolli")

Marianna Nicolosi-Asmundo (Università di Catania)

Marina Paino (Università di Catania)

Giuseppe Palazzolo (Università di Catania)

Jonathan Prag (University of Oxford Merton College)

Daria Spampinato (CNR Istituto di Scienze e Tecnologie della Cognizione)

Rachele Sprugnoli (Università di Parma)

Francesco Stella (Università di Siena)

## **Segreteria scientifica / Scientific Secretariat**

Liborio Barbarino (Università di Catania)

Denise Bruno (Università di Catania)

Giulia Cacciatore (Università di Catania)

Giuseppe Canzoneri (Università di Catania)

Elisa Conti (Università di Catania)

Milena Giuffrida (Università di Catania)

Miryam Grasso (Università di Catania)

Francesca Prado (Università di Catania)

Emilio M. Sanfilippo (CNR Istituto di Scienze e Tecnologie della Cognizione)

Eliana Vitale (Università di Catania)

Alessandro Zammataro (Università di Catania)

Comunicazione istituzionale: Claudia Cantale (Università di Catania) e Area Per la Comunicazione dell'Università di Catania (ACOM).

*Institutional communication: Claudia Cantale (University of Catania) and the Area for Communication of the University of Catania (ACOM)*

Supporto tecnico: Rosario Agrò, Area della Terza Missione dell'Università di Catania, per la consulenza e la progettazione grafica dei materiali informativi del convegno.

*Technical support: Rosario Agrò, Third Mission Area of the University of Catania, for advice and graphic design of the conference information materials.*

## **Enti organizzatori / Organisers**

AIUCD; Università di Catania: Dipartimento di Scienze Umanistiche; CNR Istituto di Scienze e Tecnologie della Cognizione; CINUM: Centro di Informatica Umanistica dell'Università di Catania.

## **Supporter**

CLARIN-IT; Neperia Group; Storage; programma Piaceri 2020-2022, Linea 1; Parmalat-Sole.

## Chair di area/ Track chair

### Le culture digitali nel Mediterraneo

Cristina Marras (CNR Istituto del Lessico intellettuale europeo e Storia delle Idee)

Paola Moscati (CNR Istituto di Scienze del Patrimonio Culturale)

### Archivi ed edizioni digitali

Christian D'Agata (Università di Catania)

Greta Franzini (Eurac Research)

### Analisi computazionale dei testi

Angelo Mario Del Grosso (CNR Istituto di Linguistica Computazionale "Antonio Zampolli")

Simone Reborra (Università di Verona)

### Ontologie e Semantic Web

Marianna Nicolosi Asmundo (Università di Catania)

Francesca Tomasi (Università di Bologna)

### Preservazione della memoria e del patrimonio digitale

Fabio Ciraci (Università del Salento)

Anna Maria Marras (Università di Torino)

## Lista dei revisori /List of reviewers

**Maristella Agosti** (Università di Padova), **Stefano Allegrezza** (Università di Bologna), **Chiara Alzetta** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Liborio Barbarino** (Università di Catania), **Nicola Barbuti** (Università di Bari Aldo Moro), **Stefano Bazzaco** (Università di Verona), **Benedetta Bessi** (Università Ca' Foscari di Venezia), **Andrea Bolioli** (ricercatore indipendente), **Paolo Bonora** (Università di Bologna), **Federico Boschetti** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Dominique Brunato** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Flavia Bruni** (Università Gabriele D'Annunzio di Chieti-Pescara), **Marina Buzzoni** (Università Ca' Foscari di Venezia), **Alberto Campagnolo** (Université Catholique de Louvain/KULeuven), **Anna Cappellotto** (Università di Verona), **Emmanuela Carbé** (Università di Siena), **Vittore Casarosa** (CNR Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" – Università di Pisa), **Fabio Ciotti** (Università di Roma "Tor Vergata"), **Fabio Ciraci** (Università del Salento), **Elisa Conti** (Università di Catania), **Salvatore Cristofaro** (CNR Istituto per il Lessico Intellettuale Europeo e Storia delle Idee), **Christian D'Agata** (Università di Catania), **Elisa D'Argenio** (HUN-REN Hungarian Research Centre for Linguistics), **Mauro De Bari** (Università di Bari Aldo Moro), **Riccardo Del Gratta** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Angelo Mario Del Grosso** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Saulo Delle Donne** (Università del Salento), **Giorgio Maria Di Nunzio** (Università di Padova), **Antonio Di Silvestro** (Università di Catania), **Filippo Diara** (Università di Torino), **Giulia Fabbris** (Università Ca' Foscari di Venezia), **Riccardo Fedriga** (Università di Bologna), **Franz Fischer** (Università Ca' Foscari di Venezia), **Greta Franzini** (Eurac Research), **Francesca Frontini** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Daniele Fusi** (Stuttgart University & VeDPH – Università Ca' Foscari di Venezia), **Carola Gatto** (Università del Salento), **Lucia Giagnolini** (Università di Bologna), **Emiliano Giovannetti** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Milena Giuffrida** (Università di Catania), **Edmondo Grassi** (Università San Raffaele di Roma), **Miryam Grasso** (Università di Catania), **Alessandro Iannella** (Università di Cagliari - Università di Pisa – Università di Torino), **Paola Italia** (Università di Bologna), **Maurizio Lana** (Università del Piemonte Orientale), **Pietro Maria Liuzzo** (Bibliotheca Hertziana), **Dominique Longrée** (Université de Liège), **Francesco Mambrini** (Università Cattolica del Sacro Cuore di Milano), **Tiziana Mancinelli** (Istituto Italiano di Studi Germanici), **Anna Maria Marras** (Università di Torino), **Cristina Marras** (CNR Istituto del Lessico intellettuale europeo e Storia delle Idee), **Federico Meschini** (Università della Toscana), **Alessio Miaschi** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Paolo Monella** (Università Sapienza di Roma), **Ouafae Nahli** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Marianna Nicolosi-Asmundo** (Università di Catania), **Giuseppe Palazzolo** (Università di Catania), **Valentina Pasqual** (Università di Bologna), **Gianluca Pavan** (Università di Roma "Tor Vergata"), **Giulia Pedonese** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Jonathan Prag** (University of Oxford Merton College), **Simone Reborra** (Università di Verona), **Giulia Renda** (Università di Bologna), **Roberto Rosselli Del Turco** (Università di Torino), **Enrica Salvatori** (Università di Pisa), **Emilio M. Sanfilippo** (CNR Istituto di Scienze e Tecnologie della Cognizione), **Eva Sassolini** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Pietro Sichera** (CNR Istituto di Scienze e Tecnologie della Cognizione), **Daniele Silvi** (Università di Roma "Tor Vergata"), **Elena Spadini** (University of Basel), **Daria Spampinato** (CNR Istituto di Scienze e Tecnologie della Cognizione), **Linda Spinazzè** (Università Ca' Foscari di Venezia), **Rachele Sprugnoli** (Università di Parma), **Francesco Stella** (Università di Siena), **Cecilia Tamagnini** (Università di Bologna), **Timothy Tambassi** (Università Ca' Foscari di Venezia), **Francesca Tomasi** (Università di Bologna), **Marco Venuti** (Università di Catania), **Fabio Vitali** (Università di Bologna).

# SOMMARIO

PREFAZIONE	I
<i>Antonio Di Silvestro, Daria Spampinato</i>	
<b>RELAZIONI DEI KEYNOTE SPEAKERS</b>	<b>VI</b>
INFORMATICA UMANISTICA, INFOCRAZIA, E INTELLIGENZE ARTIFICIALI	VII
<i>Giuseppe Savoca</i>	
THE MEDIEVAL MEDITERRANEAN IN...DATA? INTERPRETATION, CONJECTURE AND DIGITAL METHODS	XIII
<i>Tara L. Andrews</i>	
<b>RELAZIONI DEGLI INVITED SPEAKERS</b>	<b>XXI</b>
RICOSTRUZIONE DEL TESTO E BANCHE DATI. LA FILOGIA DIGITALE ALLA PROVA DELL'ESEGESI ANTICA DELLA COMMEDIA	XXII
<i>Vittorio Celotto, Andrea Mazzucchi</i>	
IL MEDITERRANEO: UN MARE DI OPPORTUNITÀ E SFIDE	XXX
<i>Salvatore Capasso</i>	
<b>MEDITERRANEO TRA TESTI E CONTESTI</b>	<b>1</b>
COMBINING GENERATIVE AI AND ARCHAEOLOGY TO BUILD DATA-DRIVEN STORIES	2
<i>Francesca Buscemi, Angelica Lo Duca</i>	
DIGITAL PRESERVATION E SOSTENIBILITÀ AMBIENTALE	9
<i>Adele Gorini</i>	
IL PROGETTO DIGITALE SU ELISA CHIMENTI (LAI-ALEEF): LA PROBLEMATICITÀ DI UN PROFILO MEDITERRANEO TRA RETI E FRONTIERE	15
<i>Ada Desideri, Bianca Vallarano</i>	
MEDITERRANEO VERDE FANTASTICO: UN REPERTORIO DIGITALE ATTRAVERSO LA COLLEZIONE BOTANICA DI ULISSE ALDROVANDI (1522-1605)	21
<i>Sara Obbiso</i>	
PER LA DIGITALIZZAZIONE DEL PATRIMONIO LINGUISTICO E CULTURALE ITALIANO IN EGITTO: I PERIODICI ITALIANI (1892- 1940)	26
<i>Wafaa El Beih</i>	
PIATTAFORME WIKI PER L'INSEGNAMENTO UMANISTICO: SPERIMENTAZIONI IN CORSO NEL LICEO DE COSMI DI PALERMO	32
<i>Antonino Fiorino, Paolo Monella, Francesca Saieva, Antonella Sorci</i>	
THE ORGANIZATION AND MANAGEMENT OF THE MAGIC PROJECT FOR ANCIENT MANUSCRIPTS DIGITIZATION: CONNECTIONS BETWEEN MEDITERRANEAN CULTURES	38
<i>Stefania Conte, Andrea Mazzucchi, Guido Russo, Augusto Tortora, Giorgia Tortora</i>	
TOWARDS A RESEMANTISATION OF THE CONCEPT OF MODELLING IN DIGITAL HUMANITIES	43
<i>Cristina Marras, Arianna Ciula, Øyvind Eide, Patrick Sahle</i>	
VALORIZZARE UN ARCHIVIO 'MEDITERRANEO': STUDI PER UN'EDIZIONE CRITICA DIGITALE DELLE OPERE DI GIOVANNI COMISSO	48
<i>Marco Borrelli</i>	
<b>ARCHIVI E MUSEI DIGITALI PER IL PATRIMONIO CULTURALE</b>	<b>55</b>
A WORKFLOW FOR GLAM METADATA CROSSWALK	56
<i>Arianna Moretti, Ivan Heibi, Silvio Peroni</i>	
DIGITALIZZAZIONE E MODELLAZIONE DELLA DRAMMATURGIA DI LEONE ALLACCI	63
<i>Luca Giovannini, Giorgia Gallucci</i>	

FROM DATA COMPLEXITY TO USER SIMPLICITY: A FRAMEWORK FOR LINKED OPEN DATA RECONCILIATION AND SERENDIPITOUS DISCOVERY	67
<i>Marco Grasso, Giulia Renda, Marilena Daquino</i>	
GIUSEPPE CHIARINI: UN'OPERA INEDITA	73
<i>Elena Almangano, Mirko Castaldi, Eleonora De Longis, Daniele Pasqualetti</i>	
HERITRACE: TRACING EVOLUTION AND BRIDGING DATA FOR STREAMLINED CURATORIAL WORK IN THE GLAM DOMAIN	78
<i>Arcangelo Massari, Silvio Peroni</i>	
LIBRI E BIBLIOTECHE TRA MUSEABILITÀ E MUSEALIZZAZIONE DIGITALE: SOGNO O REALTÀ?	84
<i>Nicola Barbuti, Mauro De Bari</i>	
LISTENING2PAINTING: AN AUDIO AUGMENTED REALITY APPROACH FOR ARTS	89
<i>Nicola Orio, Daniel Zilio, Andrea Micheletti</i>	
LUOGHI COMUNI: METODI E STRATEGIE DI SVILUPPO SOFTWARE IN AMBITO GLAM, DALLE VOCI DI AUTORITÀ ALL'ESPLORAZIONE CARTOGRAFICA	92
<i>Herbert Natta, Gianluca Rossi, Roberta Maggi</i>	
MESSAGGISTICA ISTANTANEA E ARCHIVI DIGITALI. QUALI SOLUZIONI? BEST PRACTICES E CONSIDERAZIONI DAL CONTESTO INTERNAZIONALE	98
<i>Alessia Del Bianco</i>	
NUOVE INTERAZIONI CON COLLEZIONI DIGITALI: L'“ARCHIVIO DIGITALE DEL CAPITOLO DI LATERZA”	104
<i>Stefania Riso, Nicola Barbuti</i>	
PRESERVARE E VALORIZZARE LA MEMORIA DI ARCHIVI STORICI DI EX-OSPEDALI PSICHIATRICI	110
<i>Grazia Serratore</i>	
PRESERVAZIONE DEL PATRIMONIO CULTURALE E CLOUD COMPUTING: CARATTERISTICHE E CRITICITÀ	117
<i>Manuela Grillo</i>	
PRESERVING CULINARY TRADITIONS. A CROWDSOURCED DIGITAL COLLECTION OF COOKBOOKS	122
<i>Giulia Renda, Giulia Manganelli, Mila Fumini, Marilena Daquino</i>	
SERIOUS GAMES E GAMIFICATION: A CHE PUNTO SONO LE ISTITUZIONI CULTURALI ITALIANE?	128
<i>Vincenzo Colaprice</i>	
THE TREE OF PHILOSOPHERS: DESIGN AND IMPLEMENTATION OF A DIGITAL RESOURCE FOR THE HISTORY OF ACADEMIC PHILOSOPHY	133
<i>Guido Bonino, Nicola Ruschena</i>	
THINKING OUTSIDE THE BLACK BOX: INSIGHTS FROM A DIGITAL EXHIBITION IN THE HUMANITIES	138
<i>Sebastian Barzaghi, Alice Bordignon, Bianca Gualandi, Silvio Peroni</i>	
<b>EDIZIONI SCIENTIFICHE DIGITALI</b>	<b>143</b>
IL PROGETTO CORR<SI>CA: EDIZIONE DIGITALE DELLA CORRISPONDENZA CANIONI	144
<i>Anna Giaufret, Beatrice Dal Bo, Elena Margherita Vercelli, Laura Bonanno</i>	
IL PROGETTO “MAXIMHUM”: ITALIA UMANISTICA E MOSCOVIA CINQUECENTESCA DIALOGANO IN DIGITALE	152
<i>Francesca Romoli, Letizia Ricci, Angelo Mario Del Grosso</i>	
L'ARCHIVIO DI GIUSEPPE FAVA: CONSERVAZIONE E VALORIZZAZIONE ATTRAVERSO IL DIGITALE	159
<i>Giuseppe Davide Di Mauro, Marzia D'Amico</i>	
L'ARCHIVIO DIGITALE DI UNA CASA EDITRICE: L'ESEMPIO DEL SAGGIATORE E DELLA SUA PRIMA PUBBLICAZIONE	165
<i>Giada Di Pino</i>	
MARCARE LA POESIA DEL NOVECENTO: UNO STUDIO PER <i>OSSI DI SEPPIA</i>	171
<i>Chiara Cauzzi, Martina Corti, Anna Guadagnoli, Maria Grazia Schiaroli</i>	
METASCRIP: A FRAMEWORK PROPOSAL FOR SCREENPLAY ENCODING	175
<i>Erica Andreose, Giorgia Crosilla, Leonardo Zilli</i>	
OPENDATA: OPENGADDA	181
<i>Eleonora Pasquale, Martina Pensalfini</i>	
PAUL KLEE, <i>TUNISREISE</i> E <i>BILDNERISCHE FORMLEHRE</i> : UN CASO STUDIO DI DISCEPT (DIGITAL SCHOLARLY EDITIONS PLATFORM AND ALIGNED TRANSLATIONS)	186
<i>Hansmichael Hohenegger, Tiziana Mancinelli, Fabio Ciotti, Federico Boschetti, Angelo Mario Del Grosso, Eleonora De Longis</i>	

PAVES-E: PER UNA HYPEREDIZIONE DELL'OPERA DI CESARE PAVESE	191
<i>Christian D'Agata, Angelo Mario Del Grosso, Laura Nay, Giuseppe Palazzolo, Antonio Sichera, Daria Spampinato</i>	
PER UN'EDIZIONE DIGITALE DI <i>SE QUESTO È UN UOMO</i>	197
<i>David Tagliacozzo</i>	
PER UN'EDIZIONE SCIENTIFICA DIGITALE DELLO <i>SPECULUM GUY OF WARWICK</i>	204
<i>Omar Khalaf, Sibilla Siano</i>	
PER UNA LETTURA ANTROPOLOGICA DI VERGA: TRA CODIFICA E GEOREFERENZIAZIONE	210
<i>Giovanna Zisa</i>	
PROGETTO DI EDIZIONE GENETICA DIGITALE DEL <i>CANZONIERE</i> MANOSCRITTO DI U. SABA (1919-20)	215
<i>Marina Buzzoni, Davide Cucurnia, Cristina Fenu, Roberto Rosselli Del Turco, Giulia Tancredi</i>	
RAPPRESENTARE LA STORIA SACRA: UN'IMPRESA IERI, UNA SFIDA OGGI. PROPOSTA DI EDIZIONE SCIENTIFICA DIGITALE DEL "COMPENDIUM" DI PIETRO DI POITIERS	221
<i>Franz Fischer, Agnese Macchiarelli</i>	
TOWARDS AN INTEGRATED DIGITAL EDITION OF THE <i>LEGES LANGOBARDORUM</i>	226
<i>Marina Buzzoni, Roberto Rosselli Del Turco</i>	
UN <i>CORPUS</i> ONLINE DELLA LETTERATURA SECONDARIA (1872- 1890) DEL VERISMO ITALIANO	232
<i>Denise Bruno, Giuseppe Canzoneri, Antonio Di Silvestro, Daria Spampinato, Alessandro Zammataro</i>	
UN PROGETTO DI EDIZIONE DIGITALE <i>IMAGE-BASED</i> DELLE <i>MERAVIGLIE D'ORIENTE</i> NEL MS COTTON VITELLIUS A.XV	240
<i>Andreea Mihaela Toma</i>	
UNA PROPOSTA DI CODIFICA IN XML/MEI PER TESTI MUSICALI AUTOGRAFI DI VINCENZO BELLINI	246
<i>Laura Mazzagufò</i>	
VERISMO DIGITALE. PER UN'EDIZIONE DIGITALE COMMENTATA DELLE OPERE DI VERGA, CAPUANA, DE ROBERTO	252
<i>Liborio Pietro Barbarino, Elisa Conti, Christian D'Agata, Miryam Grasso, Ninna Maria Lucia Martines, Eliana Vitale</i>	
VERSO L'HYPEREDIZIONE. LO SVILUPPO DI PIRANDELLO NAZIONALE TRA DIDATTICA E RICERCA	260
<i>Milena Giuffrida, Christian D'Agata, Giulia Cacciatore, Fabrizio Lo Presti</i>	
VOCI DALL'INFERNO: DANTE PER DIRE IL LAGER - DIGITALIZZARE E STUDIARE LE TESTIMONIANZE	267
<i>Angelo Mario Del Grosso, Marina Riccucci, Elvira Mercatanti</i>	
<b>DIZIONARI E DIGITALIZZAZIONE DI BANCHE DATI</b>	<b>274</b>
IL VIVER (VOCABOLARIO DELL'ITALIANO VERISTA)	275
<i>Gabriella Alfieri, Marco Biffi, Stephanie Cerruto, Giovanni Salucci</i>	
L'INFORMATIZZAZIONE DEL GDLI: RISULTATI, PROSPETTIVE, SFIDE FUTURE	281
<i>Eva Sassolini, Sebastiana Cucurullo, Marco Biffi</i>	
LA DIGITALIZZAZIONE DEL DIZIONARIO LATINO LANA 1978	287
<i>Francesca Michelone</i>	
LEXICAD: PIATTAFORMA LESSICOGRAFICA DIGITALE PER L'ITALIANO DELLE ORIGINI	293
<i>Salvatore Arcidiacono, Antonella Zammataro</i>	
STRATEGIE PER LA CREAZIONE E LA CONDIVISIONE DI UNA COLLEZIONE DIGITALE DI TESTI GRECO-LATINI	298
<i>Vincenzo Ortoleva, Maria Rosaria Petringa, Salvatore Cammisuli, Mariarosaria Villareale</i>	
THE CORR<SI>CA PROJECT: ENHANCING AND "QUERYING" THE CANIONI FAMILY CORRESPONDENCE	303
<i>Tiziana Pasciuto, Selenia Anastasi, Daniele Zolezzi, Simonetta Acacia, Giada D'Ippolito, Chiara Storace, Maria Tolaini</i>	
XML/TEI E DIZIONARI <i>BORN-DIGITAL</i> : UNA PROPOSTA PER LE RISORSE LESSICOGRAFICHE DELLA RETE LEXICAD/PLUTO	310
<i>Giuseppe Leonardo Zappalà</i>	
<b>ANALISI COMPUTAZIONALE, INTELLIGENZA ARTIFICIALE E LINGUISTICA</b>	<b>318</b>
ANALISI COMPUTAZIONALE DEI REPORT DI SOSTENIBILITÀ: LA VAGHEZZA COME STRATEGIA DI GREENWASHING	319
<i>Erica Cutuli</i>	
ANALISI STILOMETRICA APPLICATA ALLE CAPACITÀ EMULATIVE DI GPT-4	325
<i>Marco De Cristofaro, Mariangela Giglio</i>	
C'È UN TESTO IN QUESTA CHAT? INTELLIGENZA ARTIFICIALE E COOPERAZIONE INTERPRETATIVA	332
<i>Daniele Silvi</i>	

GENERE E GEOPOLITICA NELLE DISCIPLINE UMANISTICHE DIGITALI IN ITALIA. LE CONFERENZE AIUCD (2012-2023)	
<i>Selenia Anastasi</i>	336
GLI LLM COME LETTORI MODELLO ARTIFICIALI	342
<i>Fabio Ciotti</i>	
I SIMILI SI ATTRAGGONO. LA VALUTAZIONE LETTERARIA SULLE PIATTAFORME DI DIGITAL SOCIAL READING	348
<i>Gabriele Vezzani, Simone Rebora, Massimo Salgaro</i>	
IL <i>DISTANT READING</i> È L'ORNITORINCO	354
<i>Pietro Mazzarisi</i>	
L'IMPIEGO DELL'INTELLIGENZA ARTIFICIALE PER LA RICOSTITUZIONE DELLE AGGREGAZIONI ARCHIVISTICHE E L'ARRICCHIMENTO DEI METADATI NEGLI ARCHIVI DIGITALI	361
<i>Stefano Allegrezza</i>	
MACCHINE PER LEGGERE: LA TEXT ANALYSIS COME STRUMENTO PER IMPARARE A LEGGERE I CLASSICI DELLA NARRATIVA... E AD AMARLI	367
<i>Fabio Ciotti</i>	
PRESERVARE LA DIVERSITÀ NELL'ERA DELL'INTELLIGENZA ARTIFICIALE: IL DILEMMA ETICO DI BIAS E DISCRIMINAZIONI NEGLI ALGORITMI	371
<i>Gianluca Pavani</i>	
<i>QUI PRO QUO?</i> DATI TESTUALI E STRUMENTI PER LA RISOLUZIONE DI COREFERENZE IN LATINO	377
<i>Roberta Grazia Leotta, Eleonora Delfino</i>	
STRUMENTI DIGITALI PER LA TRASCRIZIONE E LA LEMMATIZZAZIONE DI TESTI IN ITALIANO ANTICO	382
<i>Emiliano Degl'Innocenti, Alessia Spadi, Federica Spinelli, Lucia Francalanci, Michela Perino, Irene Falini, Francesco Coradeschi, Francesco Pinna</i>	
TESTI ALLOGRAFICI: CONTATTI TRA LINGUE E SCRITTURE DEL MEDITERRANEO	388
<i>Antonio Pagliara, Federico Boschetti, Daniele Baglioni</i>	
THE DARK MIRROR OF ARTIFICIAL INTELLIGENCE: HOW AI AFFECTS CLIMATE CHANGE	393
<i>Mauro De Bari</i>	
UN SISTEMA DI CLASSIFICAZIONE AUTOMATICA DI IMMAGINI RELATIVE A MATERIALI LIBRARI ANTICHI E MODERNI	398
<i>Nicola Barbuti, Tommaso Caldarola</i>	
UNCOVERING THE SPREAD OF LEXICAL INNOVATION IN ITALIAN TWEETS	405
<i>Greta Franzini, Paolo Brasolin, Stefania Spina</i>	
<b>ORGANIZZAZIONE DELLA CONOSCENZA CON SEMANTIC WEB</b>	<b>410</b>
AFFINARE IL CONTESTO: ESTRAZIONE DI INFORMAZIONI STRUTTURATE PER L'ARRICCHIMENTO DEI CONTESTI ARCHIVISTICI	411
<i>Lucia Giagnolini, Paolo Bonora, Francesca Tomasi</i>	
CLEF 2.0. SOLUZIONI PER LA CATALOGAZIONE NATIVA LINKED DATA DEL PATRIMONIO DIGITALE CULTURALE ITALIANO	417
<i>Marilena Daquino, Laurent Fintoni, Sebastiano Giacomini, Francesca Tomasi</i>	
L'ONTOLOGIA BIGRAFO: VERSO UN MODELLO SEMANTICO PER L'OPERA DI FRANCO FORTINI	423
<i>Laura Antonietti, Emilio M. Sanfilippo, Emmanuela Carbé</i>	
LOST IN DATIFICATION? THE JOURNEY OF DATA FROM THE PRIMARY SOURCE TO THE FINAL INTERPRETATION	429
<i>Enrica Bruno, Sofia Baroncini, Francesca Tomasi</i>	
ORBIS DIOECESIVM. CREATING AUTHORITY DATA ON THE LEGAL-HISTORICAL CHANGES OF CATHOLIC DIOCESES	435
<i>Benedetta Albani, Rowan Dorin, Yohan Park</i>	
PER L'INTEROPERABILITÀ E LA SOSTENIBILITÀ DELLE RISORSE DIGITALI DANTESCHE: IL PROGETTO LIDA	441
<i>Cesare Concordia, Gaia Tomazzoli, Nicola Aloia, Carlo Meghini, Luca Trupiano</i>	
PER UN'ANALISI DEI PERSONAGGI TRA LETTERATURA, FILOSOFIA E ONTOLOGIA APPLICATA	448
<i>Emilio M. Sanfilippo, Gaia Tomazzoli, Michele Paolini Paoletti, Jansan Favazzo, Roberta Ferrario</i>	
REPRESENTING TEXTS AS LOD: A SYSTEMATIC LITERATURE REVIEW	455
<i>Michela Bandini, Valeria Quochi</i>	
STRUCTURING AUTHENTICITY ASSESSMENTS ON HISTORICAL DOCUMENTS USING LLMs	463
<i>Andrea Schimmenti, Valentina Pasqual, Francesca Tomasi, Fabio Vitali, Marieke van Erp</i>	
VALORIZZAZIONE DI REGISTRAZIONI STORICHE DI CANTO LIRICO NEL WEB SEMANTICO	469
<i>Marcello Ranieri, Angelo Pompilio</i>	



<b>PUBLIC HISTORY E ARCHEOLOGIA DIGITALE</b>	<b>473</b>
ARCHEOLOGIA E RILIEVO 3D: UNA RIFLESSIONE SULLE METODOLOGIE. DUE CASI STUDIO DI AREA MEDITERRANEA	474
<i>Graziana D'Agostino, Pietro Maria Militello</i>	
DAL CATASTO BORBONICO ALLA GENOMICA. PIATTAFORME DIGITALI E INTERDISCIPLINARITÀ: IL PROGETTO “WE ARE WHAT THEY WERE” DI RIPOSTO	481
<i>Salvatore Spina</i>	
DRAWING HISTORY FROM THE CODICE PELAVICINO DOCUMENTS: GRAPH VISUALIZATION FOR HUMAN RESEARCHERS	489
<i>Natthida Wiwatwicha</i>	
GESTIONE INFORMATICA DELLA DOCUMENTAZIONE ARCHEOLOGICA "MINORE". METODOLOGIE E APPLICAZIONI NELL'AMBITO DEL PROGETTO STORAGE	494
<i>Marianna Figuera, Erica Platania</i>	
GESTIRE L'IMMATERIALE. CONTESTI SENSORIALI A SERVIZIO DEL PATRIMONIO ARCHEOLOGICO	501
<i>Serena D'Amico</i>	
GODSCAPES: MODELING SECOND MILLENNIUM BCE POLYTHEISMS IN THE EASTERN MEDITERRANEAN THANKS TO THE STORAGE PROJECT	507
<i>Nicola Laneri, Marianna Nicolosi-Asmundo, Daniele Francesco Santamaria, Chiara Pappalardo</i>	
IL PROGETTO STORAGE: DAI DATI AL WEB	512
<i>Simone Faro, Pietro Maria Militello, Marianna Nicolosi-Asmundo</i>	
LA MODELLAZIONE 3D AL SERVIZIO DELL'ARCHEOLOGIA: NUOVE PROSPETTIVE PER L'APPLICAZIONE AD EDIFICI MULTIPIANO DI ETÀ PROTOSTORICA	517
<i>Dario Puglisi, Marco Chiricallo</i>	
MLS CON SENSORE LIDAR APPLE PER LO SCAVO ARCHEOLOGICO: APPLICAZIONI PRATICHE	523
<i>Luigi M. Calì, Antonello Fino, Gian Michele Gerogiannis</i>	
ODONIMI D'ITALIA E DIGITAL PUBLIC HISTORY: LE PROBLEMATICHE DI UNA SCHEDATURA PARTECIPATA	528
<i>Enrica Salvatori, Vittore Casarosa, Riccardo Chiari</i>	
OPENSTREETMAP: UNO STRUMENTO E UNO SPAZIO PER LA DIGITAL PUBLIC HISTORY?	535
<i>Camilla Zucchi</i>	
PRESERVARE LA MEMORIA: IL PROGETTO STORAGE E L'ARCHIVIO DELL'EX ISTITUTO DI ARCHEOLOGIA	539
<i>Giovanni Fragalà, Pietro Maria Militello</i>	
UN ATLANTE DIGITALE PER LA STORIA MARITTIMA DEL REGNO DI SARDEGNA	545
<i>Giampaolo Salice</i>	
UN FUTURO PER LA MEMORIA. STRUMENTI, MODELLI E SINERGIE PER L'INTEGRAZIONE DEI DATI NEL PORTALE DELLE FONTI PER LA STORIA DELLA REPUBBLICA ITALIANA	549
<i>Michela Tardella, Roberta Maggi, Giorgia Lodi, Riccardo Albertoni, Herbert Natta, Gianluca Rossi, Tiziana Pasciuto, Paola Ciandrini, Luca Sinopoli, Maria Teresa Artese, Isabella Gagliardi, Eleonora Lattanzi, Sara Ventroni, Elisa Tizzoni, Alessandro Russo, Margherita Porena</i>	
<b>INFRASTRUTTURE PER LA RICERCA UMANISTICA</b>	<b>555</b>
COPHI EDITOR: FROM GREEKSCHOOLS TO THE PROJECTS MULTIVERSE	556
<i>Simone Zenzaro</i>	
DIGITAL HUMANITIES AND HERITAGE SCIENCE: MOVING FROM LANDSCAPING TO A DYNAMIC RESEARCH OBSERVATORY IN AN OPEN SCIENCE CLOUD	559
<i>Roberta Bianca Luzietti, Alessia Spadi, Nicola Giampietro, Giacomo Mancuso, Alessandra Caravale, Antonio D'Eredità, Marta Caradonna, Paola Moscati, Valeria Quochi, Monica Monachini, Emiliano Degl'Innocenti</i>	
FUNZIONI E SOSTENIBILITÀ DI UNA PIATTAFORMA DIGITALE PER LE LINGUE ARCAICHE	566
<i>Michele Mallia, Riccardo Del Gratta, Valeria Quochi</i>	
INFRASTRUTTURE DI RICERCA COME STRUMENTI DI “INTERCULTURALITÀ DIGITALE”	572
<i>Salvatore Cristofaro, Vittoria Fabiani, Cristina Marras, Enrico Pasini, Pietro Sichera, Mingyang Yu</i>	
MATERIALI DIDATTICI COME OGGETTI DIGITALI FAIR: UNA METODOLOGIA CONDIVISA PER LA FORMAZIONE IN H2IOSC	577
<i>Giulia Pedonese, Francesca Frontini, Roberta Ottaviani, Federico Boschetti, Alessia Spadi, Lucia Francalanci, Alessia Scognamiglio, Pietro Restaneo, Antonina Chaban, Jana Striova, Laura Benassi</i>	
RETHINKING SCHOLARLY DIGITAL OBJECTS AS CULTURAL HERITAGE: THE KNOT PROJECT	582
<i>Laurent Fintoni</i>	

THE ATLAS: A KNOWLEDGE GRAPH OF DIGITAL SCHOLARLY RESEARCH ON ITALIAN CULTURAL HERITAGE	588
<i>Marilena Daquino, Alessia Bardi, Marina Buzzoni, Riccardo Del Gratta, Angelo Mario Del Grosso, Franz Fischer, Francesca Tomasi, Roberto Rosselli Del Turco</i>	
INDICE DEGLI AUTORI	593

# Per l'interoperabilità e la sostenibilità delle risorse digitali dantesche: il progetto LiDa

Cesare Concordia<sup>1</sup>, Gaia Tomazzoli<sup>2</sup>, Nicola Aloia<sup>3</sup>, Carlo Meghini<sup>4</sup>, Luca Trupiano<sup>5</sup>

<sup>1</sup>CNR Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo", Italia - cesare.concordia@isti.cnr.it

<sup>2</sup>Sapienza Università di Roma, Italia - gaia.tomazzoli@uniroma1.it

<sup>3</sup>CNR Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo", Italia - nicola.aloia@isti.cnr.it

<sup>4</sup>CNR Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo", Italia - carlo.meghini@isti.cnr.it

<sup>5</sup>CNR Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo", Italia - luca.trupiano@isti.cnr.it

## ABSTRACT

In questo contributo presentiamo LiDa (Linking Dante), un progetto che mira a promuovere una maggiore interoperabilità e sostenibilità dei dati sulle opere dantesche raccolti in progetti precedenti tramite l'utilizzo di tecnologie e linguaggi del Web Semantico. Dopo un'introduzione in cui definiamo gli scopi del progetto (§1), facciamo cenno ai limiti con cui si scontra l'annotazione in dei dati linguistici e alle potenzialità di una loro implementazione in RDF (§2); introduciamo poi gli elementi fondamentali della nuova ontologia che abbiamo elaborato a tal scopo (§3), e descriviamo la procedura con cui abbiamo realizzato un nuovo grafo di conoscenza che colleghi il testo della *Commedia* alle risorse linguistiche che lo descrivono (§4); infine, presentiamo le modalità di navigazione e interrogazione di tale grafo (§5) e riflettiamo sui risultati raggiunti e sugli sviluppi futuri (§6).

## PAROLE CHIAVE

Ontologie; web semantico; Linguistic Linked Open Data; linguistica italiana; letteratura italiana.

## 1. INTRODUZIONE

Linking Dante (LiDa) è un progetto che ha l'obiettivo di digitalizzare le opere dantesche e la conoscenza a esse relativa collegando precedenti progetti e implementando nuove funzionalità, per mettere a disposizione dell'utente una piattaforma integrata che renda possibili diverse opzioni di navigazione e interrogazione. La biblioteca digitale di LiDa è basata sulle tecnologie e sui linguaggi del Semantic Web: rispetta i principi FAIR (Findable, Accessible, Interoperable, Reusable) e si ispira alle ontologie di riferimento per la rappresentazione dei dati linguistici e del patrimonio culturale; il testo di Dante diventa così un nodo inserito in una rete dove sono connesse varie risorse che permettono di approfondire le sue caratteristiche linguistiche, semantiche e intertestuali. Per raggiungere questo scopo, LiDa ha tradotto la conoscenza relativa al testo dantesco che si trova attualmente disseminata in diversi strumenti digitali e cartacei in un grafo strutturato secondo una logica calcolabile. In particolare, LiDa integra la conoscenza raccolta dai progetti DanteSearch [9], DanteSources [1], Hypermedia Dante Network [2] e MONT [5]; in questo contributo ci concentriamo sul dataset di DanteSearch (§2), che abbiamo modellato tramite una nuova ontologia (§3) e collegato al testo della *Commedia* grazie a un algoritmo (§4), rendendolo navigabile e permettendo interrogazioni complesse (§5).

## 2. DA XML A RDF

Gli studi danteschi sono da diversi decenni all'avanguardia nel campo delle digital humanities (per una rassegna cf. [4]); uno dei progetti più pionieristici in tal senso è stato DanteSearch, una piattaforma digitale che permette la navigazione e l'interrogazione dell'intero corpus delle opere volgari e latine di Dante, che sono state lemmatizzate e annotate per descriverne la morfologia e la sintassi. Il dataset di DanteSearch è costituito da un insieme di file XML annotati secondo la codifica TEI<sup>1</sup>; ciascun file riporta una sola tipologia di annotazione: sintattica oppure morfologica; all'interno dell'annotazione sintattica sono incluse anche alcune informazioni sui dialoghi della *Commedia* (il tipo di discorso – diretto, indiretto o pensato – e il locutore). Una delle funzionalità mancanti nel sistema DanteSearch, discussa dettagliatamente dal coordinatore del progetto [9: 609], è la possibilità di interrogare i testi del corpus definendo contemporaneamente filtri sulla sintassi e sulla morfologia, e questo perché i file con annotazioni di tipo diverso hanno strutture XML/TEI non omogenee tra loro (vd. Fig. 3). Per implementare questa funzionalità in DanteSearch si potrebbero prospettare diverse soluzioni: si potrebbe ad esempio definire una nuova struttura XML/TEI, molto complessa, e mappare in essa le strutture delle diverse annotazioni, oppure creare un indice esterno con un tool specifico, o ancora implementare un modulo software che esegua l'integrazione a runtime, elaborando il risultato di query separate. Ciascuna di queste soluzioni rimane

<sup>1</sup> <https://dantesearch.dantenetwork.it/download.jsp>

problematica a causa del formato di codifica adottato: XML nasce come formato per la serializzazione dei dati, cioè per la codifica delle informazioni, e per le sue caratteristiche permette di rappresentare facilmente documenti con strutture ad albero, ma richiede soluzioni specifiche per documenti in cui le associazioni tra le parti siano più complesse, come nel caso di un testo annotato. Una delle soluzioni più usate per annotare testi è la codifica XML/TEI, considerata lo «standard internazionale [...] imprescindibile per qualunque impresa di filologia digitale» [9: 586] e adottata, come si diceva, in DanteSearch. Benché le linee guida della TEI definiscano un modulo lightweight per l'annotazione morfologica o sintattica<sup>2</sup> – e non siano dunque incompatibili con l'annotazione di un testo dalla struttura più complessa – i problemi a cui abbiamo accennato ne risultano solo parzialmente risolti, come dimostra il fatto che gli annotatori di DanteSearch hanno elaborato un modello di codifica ad hoc, limitando l'interoperabilità della marcatura con altri corpora.

In anni recenti si sono diffusi progetti che mirano invece a favorire l'interoperabilità tra grandi corpora di testi, e dunque la produzione di Linguistic Linked Open Data (LLOD), grazie ad annotazioni linguistiche fondate su ontologie come Ontolex<sup>3</sup>, elaborate in una prospettiva che non guarda primariamente al testo, ma al linguaggio più in generale, e dunque più attenta alle sue varie componenti e alle relazioni tra queste. Un'iniziativa importante in questa direzione è LiLa (Linking Latin), che si propone di connettere varie risorse linguistiche relative al latino per renderle interoperabili, e che comprende corpora annotati, lemma-banks, tree-banks per l'annotazione sintattica e altri strumenti per il NLP. L'elemento centrale dell'architettura di LiLa è però il lemma, il che permette un buon compromesso tra fattibilità e granularità [6: 75]; il nostro approccio, invece, assegna una maggior centralità al testo, inteso come stratificazione di strutture di significato che intrattengono diverse relazioni paradigmatiche e sintagmatiche.

Il progetto LiDa ha tra i suoi obiettivi quello di rendere le opere dantesche annotate interoperabili con altri corpora. Per fare questo abbiamo in primo luogo elaborato un'ontologia che descrive il testo letterario nella sua complessa relazione di strutture di significato relative rispettivamente alla partizione dell'opera, alla morfologia e alla sintassi, e che permette di connettere ogni elemento di tali strutture a modelli e vocabolari standard. Per rappresentare le entità individuate abbiamo scelto di usare RDF, un formalismo per la definizione di modelli di dati basato sui principi del Semantic Web, che permette di creare grafi di conoscenza (knowledge graphs), cioè strutture dati in cui le informazioni sono organizzate secondo regole definite da ontologie formali, e sulle quali è possibile applicare ragionatori automatici per inferire nuova conoscenza. Sui grafi di conoscenza è possibile fare ricerche semantiche e tutte le entità sono identificate da un IRI, che le rende disponibili per essere referenziate, e dunque riutilizzate, come LLOD.

### 3. CONCETTI E ONTOLOGIE

Il progetto LiDa, nello specifico, si basa su un'ontologia applicativa, espressa in OWL 2 DL e chiamata Ontology of Literary Resources (OLiRes), che a sua volta si fonda sul CIDOC CRM<sup>4</sup> e su Ontolex. OLiRes rappresenta il testo della *Commedia* di Dante secondo la sua struttura testuale, morfologica e sintattica, ed è interoperabile con le ontologie HDN (cf. [2]), MONT (cf. [5]), ORL (Ontologia delle Risorse Lessicali) e SyntIt (la nostra ontologia per la sintassi dell'italiano), che integrano ulteriore conoscenza rispettivamente sulla dimensione intertestuale codificata dal secolare commento, sulle metafore e su lessico, morfologia e sintassi della *Commedia*. OLiRes rappresenta la conoscenza sul testo delle opere letterarie a tre livelli: 1. il *livello dell'occorrenza*, che rappresenta le caratteristiche del testo nella sua concreta realizzazione, compresa la sua struttura (nel caso della *Commedia*, per esempio, la sua divisione in cantiche e canti); 2. il *livello linguistico*, che codifica gli aspetti linguistici del testo, come la sua morfologia e la sua sintassi; 3. il *livello concettuale*, che rappresenta la componente semantica e concettuale del testo, comprese le relazioni tra le forme che occorrono nel testo e i loro significati o tra il veicolo e il tenore di una metafora.

Questi tre livelli sono legati tra di loro da relazioni di *istanziamento* o di *occorrenza*: ogni entità linguistica è un'istanza della corrispondente entità concettuale e si manifesta in un'occorrenza; così, per fare un esempio, la forma 'cammin' è un'istanza del lemma 'cammino', che a sua volta rimanda al concetto di 'cammino', e occorre nel primo verso del primo canto della *Commedia*, in quarta posizione («Nel mezzo del *cammin* di nostra vita», *Inf.* I, 1). Le entità che fanno parte di uno stesso livello – dell'occorrenza, linguistico o concettuale – sono legate tra di loro da una relazione di *composizione ordinata*: una cantica è composta da una sequenza ordinata di canti, un periodo è composto da una sequenza ordinata di frasi. La relazione di composizione si lega a quella di *localizzazione*: per rappresentare le diverse entità che compongono il testo e per permetterne il reperimento tramite interrogazioni, l'ontologia rappresenta il frammento (o i frammenti) di testo in cui occorre ciascuna entità, specificandone le coordinate di inizio e fine all'interno del testo dell'opera, visto come uno spazio monodimensionale continuo e totalmente ordinato, pur con alcuni distinguo. Una determinata forma flessa, infatti,

<sup>2</sup> <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/AI.html#AILALW>

<sup>3</sup> [https://www.w3.org/community/ontolex/wiki/Main\\_Page](https://www.w3.org/community/ontolex/wiki/Main_Page)

<sup>4</sup> <https://www.cidoc-crm.org/>

occorre sempre in un unico frammento di lunghezza 1, un determinato periodo occorre sempre in un frammento di lunghezza (almeno) 1, mentre una frase può occorrere in più frammenti se nel periodo si trova intervallata da altre frasi. All'interno della nostra ontologia, ogni entità, qualunque sia il livello a cui appartiene, è rappresentata dalla classe `olires:Entity`, sottoclasse di `owl:Thing`; poiché nello standard che abbiamo preso a riferimento, il CIDOC CRM, esiste solo la classe `ecrm:E33_Linguistic_Object`<sup>5</sup>, che non distingue tra entità linguistiche e frammenti testuali, abbiamo introdotto le due sottoclassi `olires:SyntacticEntity`, avente come istanze le entità linguistiche, e `olires:TextFragment`, avente come istanze le occorrenze (vd. Fig. 2). Per rappresentare le proprietà di istanziazione e occorrenza sopra descritte abbiamo introdotto le proprietà `olires:instanceOf`, che lega ogni istanza del livello linguistico alla corrispondente istanza del livello concettuale, e `olires:occurrenceOf`, che lega ogni occorrenza alla corrispondente istanza del livello linguistico; non è stato possibile ricondurre queste due proprietà ad alcuna proprietà del CRM. Per rappresentare la proprietà di composizione, invece, abbiamo usato la proprietà `ecrm:P148_has_component`, che lega le istanze di `ecrm:E89_Propositional_Object` tra di loro e si applica quindi sia alle entità linguistiche che alle occorrenze, essendo entrambe istanze di `ecrm:E33_Linguistic_Object`, che è sottoclasse di `ecrm:E89_Propositional_Object`.

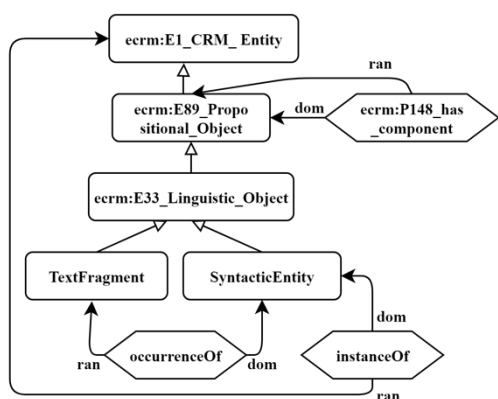


Figura 1. Entità linguistiche e testo

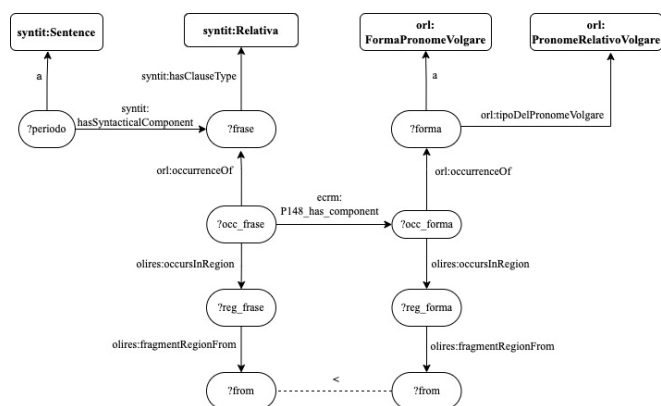


Figura 2. Es. di filtro SPARQL sul grafo della Commedia

Tuttavia, la proprietà `ecrm:P148_has_component` non può essere usata in alcun assioma che vincoli la proprietà a specifiche classi (ad esempio, l'assioma che stabilisce che le componenti di un'entità sintattica sono solo entità sintattiche, e viceversa) perché è dichiarata transitiva, quindi è una proprietà *composite*<sup>6</sup> su cui non possono essere espressi assiomi di cardinalità, pena la violazione della *Restriction on Simple Roles*. Per questo abbiamo introdotto quattro sottoproprietà intransitive di `ecrm:P148_has_component`: `olires:hasStructuralComponent` per la composizione delle entità strutturali; `olires:hasSyntacticalComponent` per la composizione delle entità lessicali e sintattiche; `hasTenor` e `hasVehicle` per la composizione delle metafore.

Come abbiamo ricordato, l'ontologia di LiDa è OLiRes, la quale si appoggia a diverse ontologie di dominio, configurandosi come un sistema composito; le due classi `olires:SyntacticEntity` e `olires:TextFragment` giocano un ruolo fondamentale nell'integrazione delle ontologie che assiomatizzano i diversi fenomeni linguistici che formano tale sistema composito. In particolare, (vd. Fig. 1) l'ontologia Ontolex, usata per la morfologia, si lega a OLiRes tramite gli assiomi che stabiliscono che `ontolex:Form`, la classe delle forme, è sottoclasse di `olires:SyntacticEntity` e che `ontolex:lexicalForm`, la proprietà che lega una forma al suo lemma, è sottoproprietà di `ecrm:P148_has_component`; per contro, la classe `olires:FormOccurrence`, che ha come istanze le occorrenze delle forme, è sottoclasse di `olires:TextFragment`; (2) l'ontologia SyntIt, usata per la sintassi dell'italiano, si lega a OLiRes tramite gli assiomi che stabiliscono che `syntit:Sentenza`, la classe dei periodi, è sottoclasse di `olires:SyntacticEntity`, mentre `syntit:SentenzaOccurrence`, la classe delle occorrenze dei periodi, è sottoclasse di `olires:TextFragment`; questa classificazione si propaga poi alle relative sottoclassi, quella delle frasi

<sup>5</sup> Secondo la Scope Note, la classe `E33_Linguistic_Object` del CIDOC CRM «comprises identifiable expressions in natural language or languages. Instances of `E33_Linguistic_Object` can be expressed in many ways: e.g., as written texts, recorded speech or sign language».

<sup>6</sup> Secondo la terminologia introdotta nella sezione 11 (Global Restrictions on Axioms in OWL 2 DL) della specifica di OWL 2 DL (<https://www.w3.org/TR/owl2-syntax/>).

(`syntit:Clause`) e relative occorrenze (`syntit:ClauseOccurrence`), quella dei sintagmi (`syntit:Syntagm`) e relative occorrenze (`syntit:SyntagmOccurrence`); (3) l'ontologia HDN, usata per le relazioni intertestuali, si lega a OLiRes tramite gli assiomi che stabiliscono che `hdn:Reference`, la classe più generale dei riferimenti intertestuali, è sottoclasse di `olires:SyntacticEntity`, mentre `hdn:ReferenceOccurrence`, la classe più generale delle occorrenze dei riferimenti intertestuali, è sottoclasse di `olires:TextFragment`; (4) l'ontologia MONT, usata per le metafore, si lega a OLiRes tramite gli assiomi che stabiliscono che `mont:LinguisticMetaphor`, la classe delle metafore linguistiche, è sottoclasse di `olires:SyntacticEntity`, mentre `mont:MetaphorOccurrence`, la classe delle occorrenze di metafore, è sottoclasse di `olires:TextFragment`. In sostanza, `olires:SyntacticEntity` e `olires:TextFragment` agiscono da elementi cardine nell'articolazione delle ontologie che concorrono a formare la base logica di LiDa, lasciando alle ontologie specifiche appena nominate il compito di rappresentare i rispettivi domini. Questo ci permette di osservare che lo sviluppo di un'ontologia applicativa non può prescindere né dal riferimento a ontologie di dominio, né dal riferimento a un'ontologia top – nel caso di LiDa il CRM e Ontolex, che forniscono i concetti di base, con in testa la classe `ecrm:E33_Linguistic_Object`, di cui `olires:SyntacticEntity` e `olires:TextFragment` sono sottoclassi.

#### 4. REALIZZAZIONE DEL GRAFO DI CONOSCENZA DELLA *COMMEDIA*

Una volta definita l'ontologia OLiRes, il grafo di conoscenza relativo è stato realizzato con un software che estrae ed elabora il testo e le annotazioni presenti nei file XML/TEI con le annotazioni morfologiche e sintattiche pubblicate in DanteSearch. Nei file con le annotazioni morfologiche della *Commedia*, uno per ogni cantica, ogni forma è annotata con un elemento `<LM>`. L'elemento ha due attributi, `@lemma` e `@catg`, i cui valori sono rispettivamente la forma canonica corrispondente alla forma annotata e una codifica delle caratteristiche grammaticali/morfologiche della forma annotata.

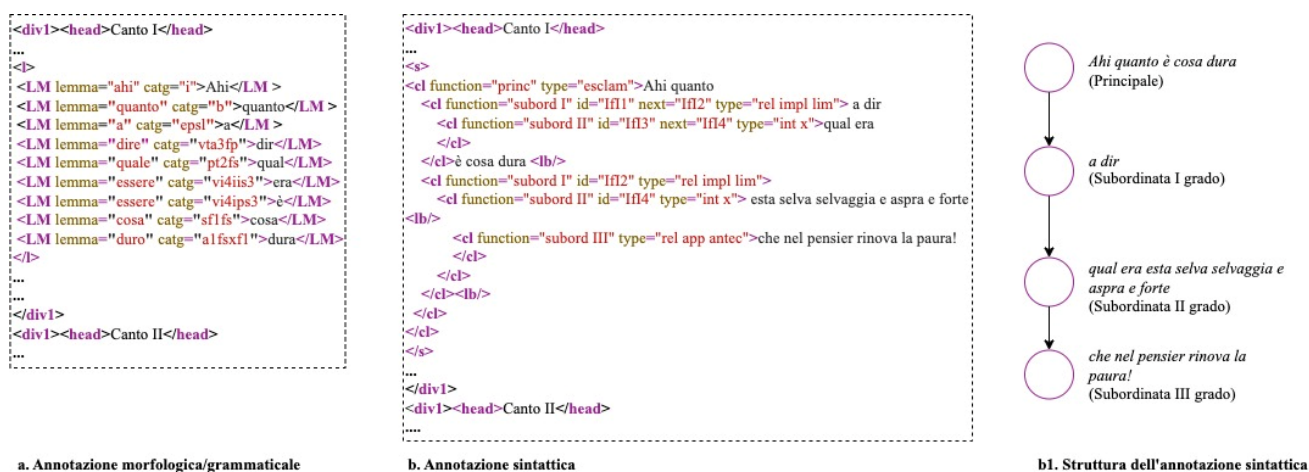


Figura 3. Esempi di annotazioni della *Commedia* in DanteSearch

Gli elementi `<div1>` delimitano le forme che appartengono a ciascun canto e gli elementi `<l>` delimitano le forme appartenenti ai versi (vd. Fig. 3a: annotazione morfologica/grammaticale di *Inf.* I, 3). Il software realizzato esegue un parsing dei tre file, processa le annotazioni e genera le entità del livello linguistico e del livello dell'occorrenza per ciascuna delle forme, inserendole nel grafo di conoscenza. L'annotazione sintattica della *Commedia* realizzata da DanteSearch è basata sull'analisi del periodo [3]: ogni canto è suddiviso in periodi, a loro volta suddivisi in frasi; i frammenti di testo che compongono una frase possono essere non consecutivi. Nel file di DanteSearch con le annotazioni sintattiche i periodi sono identificati dall'elemento TEI `<s>` (s-unit) e le frasi dall'elemento `<cl>` (clause); entrambi gli elementi possono avere l'attributo `@id` che identifica l'annotazione e l'attributo `@next` (di tipo IDREF), utilizzato per collegare in un'unica aggregazione elementi che annotano frammenti di testo non consecutivi. Il tipo sintattico<sup>7</sup> e la funzione sintattica<sup>8</sup> di ciascuna frase sono riportate negli attributi `@type` e `@function` dell'elemento `<cl>` che la annota. Le informazioni sui gradi di subordinazione tra le frasi di un periodo sono codificate sia nelle annotazioni sia nell'annidamento degli elementi

<sup>7</sup> Etichette sintetiche che rappresentano categorie desunte dalla Grande Grammatica Italiana di Consultazione [8].

<sup>8</sup> Possibili valori sono: principale, coordinata a una principale, subordinata (I-VII grado), coordinata a una subordinata (I-V grado), parentetica, coordinata a una parentetica, pseudo-coordinata.

<c1> (vd. Fig. 3b: annotazione sintattica di *Inf.* I, 3-6). Di seguito l’algoritmo per la creazione delle entità del grafo di conoscenza sintattico:

- Trasforma il file con le annotazioni sintattiche in una struttura dati che rappresenta un albero orientato: crea un nodo per ogni elemento TEI di annotazione; fra due nodi c’è un arco se i corrispondenti elementi sono in relazione parent-child (un livello di annidamento) nella struttura del file XML/TEI. Gli elementi che annotano una frase composta da frammenti di testo non consecutivi sono raggruppati in un unico nodo (vd. Fig. 3b1: parte dell’albero relativo al periodo di *Inf.* I, 3-6). La struttura dati ottenuta rappresenta un albero orientato, ed è possibile usare degli algoritmi di visita dell’albero per verificare la correttezza del parsing come pure per ottenere informazioni da aggiungere al grafo di conoscenza. Ad esempio, controllando la matrice delle adiacenze delle funzioni sintattiche delle frasi ottenuta visitando l’albero (vd. Fig. 4) sono state rilevate e corrette delle imprecisioni nelle annotazioni.
- Crea le triple delle entità linguistiche sintattiche visitando i sottoalberi che rappresentano i periodi. Parallelamente genera anche le triple del livello dell’occorrenza: per ciascun periodo/frase calcola i frammenti di testo in cui occorre e costruisce le entità del grafo di conoscenza che le rappresentano.
- Utilizzando le entità del grafo generate nella fase precedente individua le occorrenze delle forme la cui posizione è compresa nei frammenti di testo (`olires:TextFragment`) di ciascun periodo e di ciascuna frase; associa tali occorrenze a quelle dei periodi/frasi (tramite `ecrm:P148_has_component`), realizzando il collegamento tra le annotazioni morfologiche e quelle sintattiche.
- Il grafo di conoscenza della *Commedia* è memorizzato in un triple-store e accessibile tramite query SPARQL. Le condizioni di ricerca possono essere create componendo, tramite gli operatori logici, condizioni morfologiche e/o sintattiche. Inoltre, usando le proprietà dell’occorrenza di una entità linguistica, è possibile definire condizioni anche sui frammenti di testo in cui questa si manifesta.

Come esempio di query consideriamo una ricerca sulla *Commedia* descritta in [9], che può essere formulata così: “cerca le frasi relative in cui sia presente un pronome relativo che non occupi la prima posizione nella frase”. In DanteSearch questa ricerca viene eseguita solo in parte con il search engine: per essere completata richiede una fase di controllo manuale del risultato ottenuto con l’engine. Nel grafo creato in LiDa, invece, è possibile eseguire la ricerca con una singola query, definendo nel filtro di ricerca le condizioni sulla posizione nel testo delle occorrenze, sia delle forme sia delle frasi (vd. Fig. 2).

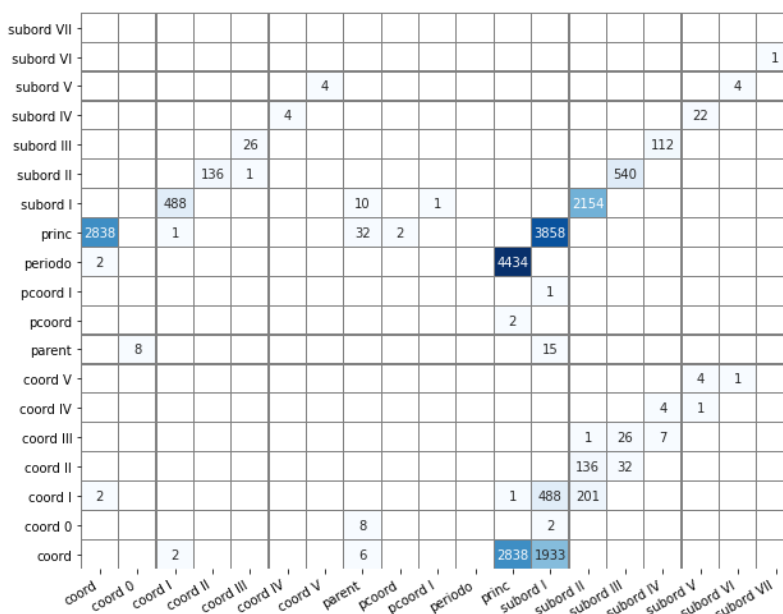


Figura 4. Adiacenze fra funzioni sintattiche

## 5. VISUALIZZAZIONE E RICERCA

Per permettere, con vari livelli di sofisticazione, l'accesso ai dati del grafo di conoscenza abbiamo realizzato un'interfaccia grafica (GUI) accessibile come Web Application<sup>9</sup>. La GUI fornisce funzionalità di navigazione nel testo della *Commedia* e due modalità di ricerca: semplice e avanzata.

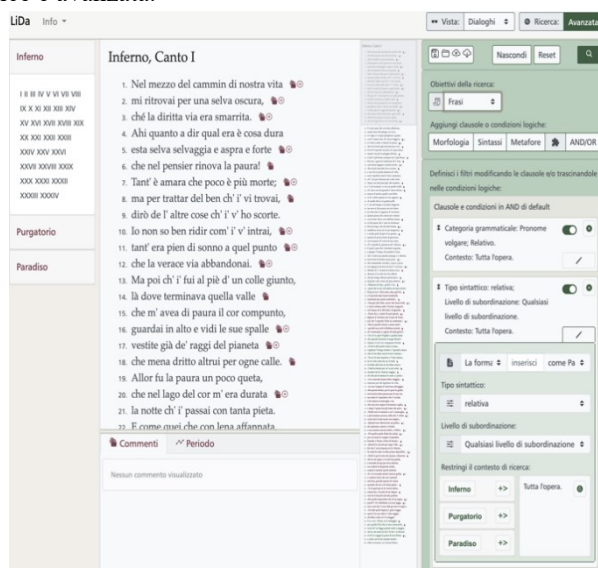


Figura 5. Visualizzazione e ricerca nel testo

La ricerca semplice permette di cercare lemmi, forme o frasi (ma anche prefissi o suffissi) sull'intera opera o su specifiche cantiche e/o canti, esprimendo opzionalmente anche altre condizioni in forma di espressioni regolari. La ricerca avanzata permette di esprimere filtri di ricerca complessi, definendo espressioni booleane i cui operandi possono essere condizioni morfologiche e/o sintattiche. Ricerca e browsing sono visualmente integrate: si può passare da una modalità all'altra senza perdita del contesto, navigando l'opera o esaminando il risultato di una ricerca. Graficamente l'interfaccia è suddivisa in tre aree sovrastate da un menù (vd. Fig. 5): a sinistra si trova un menù di navigazione tra le cantiche e i canti o tra i risultati di una ricerca; l'area di destra presenta gli strumenti per definire le query; l'area centrale contiene la visualizzazione dei dati: vi compaiono il testo del canto selezionato e altre informazioni. Quest'ultima area di visualizzazione è suddivisa orizzontalmente in due zone: nella zona superiore viene visualizzato il canto, con diverse tipologie di visualizzazione, chiamate viste:

- Vista Forme: posizionando il mouse su una forma, vengono visualizzate le informazioni morfologiche e lessicali pertinenti.
- Vista Periodi: tramite simboli e colori vengono delimitati sia i periodi in cui il canto è suddiviso, sia le frasi che compongono ciascun periodo; posizionando il mouse su una frase viene mostrata la sua funzione sintattica.
- Vista Dialoghi: vengono evidenziati i dialoghi (o discorsi) presenti nel canto; per ciascun discorso è possibile visualizzare il parlante e il tipo di discorso (diretto, riportato, pensato).

Nella parte inferiore dell'area di visualizzazione è possibile vedere i commenti relativi alle fonti primarie raccolte dal progetto HDN, facendo un clic sull'icona che compare accanto ai versi per cui esiste almeno un commento, oppure, con un doppio clic su una forma, la rappresentazione grafica, con nodi e archi, della struttura del periodo in cui essa compare.

## 6. CONCLUSIONI

L'attività di LiDa è ancora in corso: restano da implementare ulteriori funzionalità di ricerca per i commenti relativi alle fonti primarie annotati dal progetto HDN; le metafore di MOnt devono ancora essere integrate nel grafo di conoscenza, e dunque anche la loro interrogazione e visualizzazione tramite la GUI dev'essere implementata. Dato lo scopo di LiDa, che è quello di rendere interoperabili tutti i dati che costituiscono la nostra conoscenza sul testo dantesco, il progetto potrà accogliere in futuro altre risorse: speriamo così di valorizzare il lavoro – passato, presente e futuro – della comunità degli studi danteschi, e di contribuire alla sostenibilità delle risorse digitali attraverso la loro riconversione al paradigma del Web Semantico, con i suoi modelli e linguaggi. L'integrazione tra i dati relativi a tanti progetti differenti non sarebbe stata possibile con una struttura rigida come quella delle annotazioni sintattiche e morfologiche basate sull'XML: sarebbe stato estremamente laborioso, se non impossibile, realizzare una struttura ad albero così stratificata da definire nel testo dantesco

<sup>9</sup> <https://lida.dantenetwork.it>



decine di migliaia di frammenti di natura e granularità diversa. Il problema non è solo quello tecnico della sovrapposizione di gerarchie (cf. [7]): un approccio che si concentri sul testo e sulla sua espressione linguistica rende evidente la varietà di strutture sintagmatiche e paradigmatiche, e dunque di unità, con cui la lingua costruisce i suoi significati. Le triple RDF costituiscono invece una struttura liquida e seguono un pattern semplice, che definisce una regione di testo e permette la sua annotazione a qualunque livello di granularità; hanno inoltre il grande vantaggio di dotare ogni singola risorsa di un IRI, che la rende disponibile per essere referenziata e dunque riutilizzata come LLOD.

## BIBLIOGRAFIA

- [1] Bartalesi, Valentina, Carlo Meghini, Daniele Metilli, Mirko Tavoni, and Paola Andriani. 'A Web Application for Exploring Primary Sources: The DanteSources Case Study'. *Digital Scholarship in the Humanities* 33, no. 4 (2018): 705–723.
- [2] Bartalesi, Valentina, Nicolò Pratelli, Carlo Meghini, Daniele Metilli, Gaia Tomazzoli, Leyla Livraghi, and Michelangelo Zaccarello. 'A Formal Representation of the Divine Comedy's Primary Sources: The Hypermedia Dante Network Ontology'. *Digital Scholarship in the Humanities* 37, no. 2 (2022): 630–643.
- [3] Gigli, Sara. 'La codifica sintattica della Commedia di Dante'. In *Sintassi dell'italiano antico e sintassi di Dante. Atti del seminario di studi (Pisa, 15-16 ottobre 2011)*, edited by Marta D'Amico, 81–96. Ghezzano: Felici, 2015.
- [4] Maselli, Matteo. 'Per una rassegna degli strumenti della critica dantesca: dai repertori testuali ai dispositivi digitali'. *Paratesto* 18 (2021): 299–337.
- [5] Meghini, Carlo, and Gaia Tomazzoli. 'Per un'ontologia delle metafore nella Commedia di Dante'. In «Per intelletto umano / e per autorità». Il contesto di formazione e diffusione culturale del poema dantesco. Atti del I Convegno HDN (Pisa-Firenze, 29-31 ottobre 2020), edited by Leyla Livraghi and Gaia Tomazzoli, 127–152. Firenze: Franco Cesati, 2022.
- [6] Pedonese, Giulia, Flavio M. Cecchini, and Marco Passarotti. 'Linking the Computational Historical Semantics Corpus to the LiLa Knowledge Base of Interoperable Linguistic Resources for Latin'. In *Language, Data and Knowledge 2023*, edited by Sara Carvalho, Anas F. Khan, Ana Ostroski Anic, Blerina Spahiu, Jorge Gracia, John P. McCrae, Dagmar Gromann, Barbara Heinisch, and Ana Salgado, 74–85. Lisboa: NOVA FCSH - CLUNL, 2023.
- [7] Renear, Allen, Elli Mylonas, and David Durand. *Refining Our Notion of What Text Really Is: The Problem of Overlapping Hierarchies*, 1993.
- [8] Renzi, Lorenzo, Matteo Salvi, and Anna Cardinaletti, eds. *Grande grammatica italiana di consultazione*. Bologna: Il Mulino, 1995.
- [9] Tavoni, Mirko. 'DanteSearch: Il corpus delle opere volgari e latine di Dante lemmatizzate con marcatura grammaticale e sintattica'. In *Lectura Dantis 2002-2009. Omaggio a Vincenzo Placella per i suoi settanta anni*, edited by Anna Cerbo, II:583–608. Napoli: Università L'Orientale, 2011.