DH CENTRO INTERDIPARTIMENTALE DI RICERCA

ASSOCIAZIONE per l'INFORMATICA UMANISTICA e la CULTURA DIGITALE

AIUCD 2022 | UNIVERSITÀ DEL SALENTO

CULTURE DIGITALI

INTERSEZIONI FILOSOFIA ARTI MEDIA

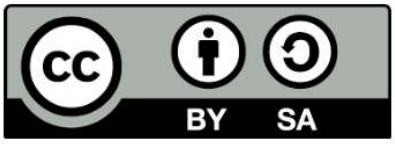
TESTO ARTI FILOSOFIA CONTENUTI INTELLIGENZA

PROCEEDINGS

ISBN 9788894253566

Copyright ©2022 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.

Fabio Ciracì, Giulia Miglietta, Carola Gatto (edd.), AIUCD 2022 - Culture digitali. Intersezioni: filosofia, arti, media. Preceedings della 11ª conferenza nazionale, Lecce, 2022. Fabio Ciracì, Giulia Miglietta, Carola Gatto (edd.), AIUCD 2022 - Digital cultures. Intersections: philosophy, arts, media. Preceedings of the 11th national conference, Lecce, 2022.

Salvo diversa indicazione, ogni link citato era attivo al 21 gennaio 2022. All links have been visited on 21th January 2022, unless otherwise indicated

Si prega di notificare all'editore ogni omissione o errore si riscontri, al fine di provvedere alla rettifica**.** Please notify the publisher of any omissions or errors found, in order to rectify them. aiucd.segreteria [at] aiucd.org

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 Scientifico di AIUCD 2022.

All the paper 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 2022 Scientific Committee.

Il programma della conferenza AIUCD 2022 è disponibile online all'indirizzo/ The AIUCD 2022 conference program is available online all'apposito indirizzo

http://aiucd2022.unisalento.it http://conference.unisalento.it/ocs/index.php/aiucd2022/index/pages/view/programma

Comitato Scientifico:

General Chair: Fabio Ciracì (Università del Salento)

Local Chair: Mario Bochicchio (Università del Salento, Università di Bari)

Membri Comitato Scientifico: Marina Buzzoni (Presidentessa AIUCD, Uni. Venezia), Federico Boschetti (Ric. ILC-CNR); Federico Meschini (Uni. Tuscia); Roberto Rosselli Del Turco (Uni Torino); Rachele Sprugnoli (Ass. Ric. Univ. Cattolica); Donato Malerba (Università Bari);

Luca Bandirali, Daniela Castaldo, Francesco Ceraolo, Stefano Cristante, Domenico M. Fazio, Manolita Francesca, Marco Mancarella, Pietro Luigi Iaia, Massimiliano Rossi, Grazia Semeraro, Franco Tommasi, Luigi Patrono (Università del Salento)

Membri del Comitato di programma*:* Mario Bochicchio (Local Chair), Luca Bandirali, Daniela Castaldo, Marco Mancarella, Pietro Luigi Iaia, Federica Epifani (Responsabile Comitato di Programma), Ilenia Colonna, Patrizia Miggiano; Carola Gatto; Giulia Miglietta; Marco Giannotta; Alessia De Blasi, Isabella Hernandez.

Direttori di Area: Luca Bandirali; Mario Bochicchio; Fabio Ciracì; Roberto Rosselli Del Turco; Marco Mancarella; Grazia Semeraro.

Segreteria del Convegno: Dott.ssa Silvia Gravili
Resp. tecnico: Carlo Tafuro; web design: Dr.ssa Paola D'Amico; comunicazione: Dr.ssa Loredana De Vitis

Enti organizzatori / Organizing institutions:

AIUCD;

Università del Salento: Centro interdipartimentale in Digital Humanities in collaborazione con i corsi di laurea in Filosofia, DAMS, Beni Culturali e Digital Humanities; ISUFI, Scuola Placetelling.

Università degli Studi Aldo Moro, Dipartimento di Informatica

Sponsor

Regione Puglia; Provincia di Lecce; Città di Lecce; CINI – Consorzio Universitario Nazionale per l'Informatica; SFI-Società Filosofica Italiana; AFC - Apulia Film Commission, Teatro Pubblico Pugliese; Argo Software.

Lista dei revisori - List of the reviewers

Agnese Addone; Tommaso Agnoloni; Luca Bandirali; Nicola Barbuti; Andrea Bellandi; Armando Bisogno; Mario Alessandro Bochicchio; Andrea Bolioli; Federico Boschetti; Dominique Brunato; Paolo Buono; Dino Buzzetti; Marina Buzzoni; Luigi Catalani; Francesco Ceraolo; Daniele Chiffi; Simona Chiodo; Fabio Ciotti; Ilenia Colonna; Christian D'Agata; Elisa D'Argenio; Riccardo De Biase; Manuela De Giorgi; Daniela De Leo; Salvatore De Masi; Pierpaolo Del Coco; Angelo Mario Del Grosso; Francesca Di Donato; Giorgio Maria Di Nunzio; Federica Epifani; Daniela Fogli; Claudio Forziati; Greta Franzini; Francesca Frontini; Emiliano Giovannetti; Edmondo Grassi; Fabiana Guernaccini; Barbara Guidi; Pietro Luigi Iaia; Benedetta Iavarone; Fahad Khan; Maurizio Lana; Angelica Lo Duca; Donato Malerba; Marco Mancarella; Tiziana Mancinelli; Chiara Mannari; Valentina Marangi; Cristina Marras; Federico Meschini; Patrizia Miggiano; Giulia Miglietta; Paolo Monella; Giovanni Morrone; Serge Noiret; Deborah Paci; Antonio Pascucci; Enrico Pasini; Luigi Patrono; Igor Pizzirusso; Simone Rebora; Massimiliano Rossi; Daniela Rotelli; Enrica Salvatori; Eva Sassolini; Daria Spampinato; Rachele Sprugnoli; Enrico Terrone; Francesca Tomasi; Francesco Tommasi; Sara Tonelli; Gennaro Vessio; Marco Salvatore Zappatore.

Indice – Table of Contents

Prefazione **I**

Sessione Arti 1 – Artemisia Gentileschi **2**
La Comédie Virtuelle **4**
Climate change & digital cultural impact, the Victoria & Albert Museum **9**
La Digitalizzazione per una fruizione del Patrimonio Culturale in sito e da remoto: il caso studio della Pala Gozzi di Tiziano **12**

Sessione Testi 1 – Claude Shannon **18**
Verso la definizione di criteri per valutare soluzioni di scholarly editing digitale: il caso d'uso GreekSchools **20**
HYLAS: A new metrical search tool for Greek and Latin poetry **26**
Stylometry and Reader Response. An Experiment with *Harry Potter* Fanfiction **30**

Sessione Intelligenza 1 – Alan M. Turing **35**
Analisi e valorizzazione del patrimonio artistico mediante Intelligenza Artificiale **37**
Un Oggetto Intelligente IoT per Migliorare le Visite Interattive di Siti di Interesse Culturale **42**
Oxoce - Motore di ricerca tematico strutturato **46**

Sessione Contenuti 1 – George Boole **49**
Funzione ecosistemica e funzione storiografica della narrazione ambientale videoludica **51**
Narrazioni mediatiche delle emergenze e processi di costruzione di *quest*: quali possibili analogie? L'incidente del "corrupted blood" in "World of Warcraft" **54**
Narrazione e interazione **59**

Sessione Testi 2 – Ada Lovelace **61**
Web e social media come nuove fonti per la storia **63**
Idee, persone, *realia*: un ambiente digitale per la Via della Seta **68**
Visualizzazione del cambiamento d'uso del maschile e femminile nei titoli occupazionali **71**
GenderedOntoComedy: Toward a Gendered Representation of Literary Characters in the Dante's Commedia **76**

Sessione Filosofia 1 – Marisa Bellisario **81**
Gli indici della prima modernità come strumento storiografico: questioni preliminari metodologiche e pratiche **83**
Indici e mappe digitali per l'iter italicum di G. W. Leibniz **86**
Ermeneutica digitale del testo filosofico. Problemi e opportunità **91**
Human Enhancement e soggetto Post-Umano alla prova delle DH: come le tecnologie digitali ci trasformano **93**

Sessione Testi 3 – Grace Murray Hopper **96**
Conservazione e fruizione di banche dati letterarie: l'archivio della poesia italiana dell'Otto/Novecento di Giuseppe Savoca **98**
«Le varianti della rosa». Per un prototipo di edizione digitale del *Nome della rosa*: interpretazione, didattica, annotazione **105**
Online lexical resources for translators: where do we stand? A (possibly meaningful) case-study **111**

Sessione Filosofia 3 – Gilbert Simondon **116**
Governare le piattaforme. Cinque proposte su pluralismo e polarizzazione online **118**
A Taxonomy of Depictive Representations: From Paintings and Sculptures to Virtual Reality **122**
Paesaggi dell'incontro mediale on-demand **126**

Sessione Contenuti 2 – Marshall McLuhan **129**
Tra Public e Digital History: la soluzione ibrida dei registri parrocchiali di Monterosso on line **131**
Una nuova mappatura digitale per i borghi delle aree interne **138**
Intelligenza artificiale e archivi audiovisivi: potenzialità e sfide del progetto "PH-Remix" **141**

Sessione Intelligenza 2 – John von Neumann **145**
Un nuovo approccio per la descrizione e gestione del patrimonio culturale digitale relativo a MAB **147**
Sulla funzionalità di un'ontologia della filosofia alto medievale. Il caso dei «Moralia in lob» di Gregorio Magno **151**
La Visualizzazione Grafica di Sensi e Relazioni Semantiche di un Lessico Computazionale della Lingua Italiana **155**

Sessione Testi 4 – Hedy Lamarr **161**
Dalla codifica alla fruizione: l'edizione digitale Bellini Digital Correspondence **163**
Dante e Petrarca allo (stesso) scrittoio. Per lo sviluppo di un'ontologia di IDP a partire dall'istanza manoscritti di Itinera **169**
Il progetto 'epistolarITA' e una proposta di applicazione di algoritmi di prossimità testuale su documenti epistolari italiani (XV-XVII s.) **172**

Sessione Testi 5 – Hélèn Metzger **177**
Visualizing the genetic process of literary works **179**
Analisi linguistica e pseudonimizzazione: strumenti e paradigmi **185**
RePIM in LOD: semantic technologies to preserve knowledge about Italian secular music and lyric poetry from the 16th-17th centuries **193**

Sessione Filosofia 2 – Giulio Cesare Vanini **196**
Computare o comporre? Riflessioni sul rapporto tra poesia e digitalità alla luce di alcune considerazioni bachelardiane **198**
Schemi, ipotesi e algoritmi. Approcci kantiani alla filosofia delle tecnologie digitali **203**
Tra chair e empiétement lo spazio topologico: contributo merleau-pontyano ai sistemi informatici **207**

Sessione Testi 6 – Katherine Johnson **210**
There and back again: what to expect in the next EVT version **212**
XML-TEI: Un modello per la filologia d'autore **218**
La svolta empirico-computazionale negli studi culturali e letterari: una nuova scienza della cultura **223**

Poster **227**
Wordforms and Meanings: un Updated Report on the LiLa Project **229**
From Close to Distant Reading. Towards the Computational Analysis of "Liber Abbaci" **232**
Citizen Humanities in Tyrol: a case study on historical newspapers **236**
Un esperimento di visualizzazione grafica della terminologia del Talmud babilonese **239**
Una edizione critica digitale per la cristianistica dell'antichità **242**
Ritmi postumani: produzione poetica e machine learning **243**
Argument-Checking: a critical Pedagogy Approach to Digital Literacy **245**
"Nostra Signora Experience": il Placetelling® in Ambiente Digitale **249**

Prefazione

L'undicesima edizione del Convegno Nazionale dell'AIUCD-Associazione di Informatica Umanistica ha per titolo *Culture digitali. Intersezioni: filosofia, arti, media*. Nel titolo è presente, in maniera esplicita, la richiesta di una riflessione, metodologica e teorica, sull'interrelazione tra tecnologie digitali, scienze dell'informazione, discipline filosofiche, mondo delle arti e *cultural studies*. Per questo motivo, il Comitato Scientifico ha individuato cinque aree funzionali alla *call for paper*, in base alle quali selezionare i contributi da presentare in occasione del convegno nazionale. Tutte le aree sono connotate da un "+D" di digitale. Tale espressione non sta a indicare un addendo esornativo e accidentale, ammiccante e modaiolo, né un supplemento alle varie discipline umanistiche con funzione integrativa o sussidiaria; essa denota invece una contaminazione, profonda e trasformativa, delle discipline umanistiche con il digitale, intendendo quest'ultimo in senso ampio, come espressione di una trasformazione scientifica e tecnologica che investe e muta la cultura e la società. Alla luce dell'informatizzazione delle conoscenze e della digitalizzazione delle pratiche, che ridiscutono limiti e poteri delle discipline istituzionali, si tratta di comprendere il nuovo ruolo delle *humanities*. Si tratta di trasformazioni che pongono problematiche inedite, ma al contempo ampliano le possibilità di indagine nei campi della tradizionale ricerca umanistica. Fedeli alla massima di Terenzio – *homo sum humani nihil a me alienum puto* – siamo convinti che una tale contaminazione individui nell'umanista un interlocutore privilegiato. Siamo cioè dell'idea che i saperi si costruiscono reciprocamente, con mutua dipendenza e in maniera interrelata, travalicando i settori scientifici e le camicie di forza delle definizioni settoriali. In questo senso il *digital humanist* rappresenta una figura capace di un supplemento di conoscenza e di una visione interdisciplinare, è abilitato a una ricerca di confine spesso difficile da caratterizzare, sia in relazione agli aspetti più squisitamente teorici dell'informatizzazione, sia in riferimento agli effetti pratici e al loro portato sociale e culturale. A questa trasformazione partecipano a pieno titolo la filosofia e le arti, come discipline chiamate a riflettere sul digitale, non solo perché da sempre si interrogano sull'uomo e sul mondo, ma anche perché ambiscono a migliorare la realtà e governare il cambiamento.

Con l'intenzione, quindi, di coinvolgere la nostra comunità a riflettere sull'intersezione fra i saperi, nell'ottica di una pluralità di culture, il Comitato Scientifico ha individuato le seguenti aree di interesse: "Testo +D", che tesaurizza ed estende la tradizione di ricerca dell'AIUCD, rivolgendosi agli studi di linguistica computazionale, edizioni digitali, progetti ipertestuali, filologia ed ecdotica digitali; "Arti +D", relativa alle tecnologie digitali per il mondo dell'arte, *digital* e *cultural heritage*; "Filosofia +D", riguardante la filosofia dell'informazione, etica ed epistemologia del digitale; "Contenuti +D", con un focus su realtà virtuale e aumentata, contenuti multimediali e transmediali, ecosistemi narrativi e spazio dei media; "Intelligenza +D", orientata alla comunicazione mediata dal computer, apprendimento digitale e sistemi di traduzione automatizzata.

Per sviluppare al meglio le aree tematiche proposte per il convegno, nei mesi di ottobre e novembre 2021 il *Centro di ricerca in Digital Humanities* dell'Università del Salento, in collaborazione con l'AIUCD, ha organizzato il ciclo di seminari "Loading AIUCD2022", a cura di Fabio Ciracì e di Patrizia Miggiano, con sette incontri in modalità telematica, in cui numerosi accademici ed esperti del settore si sono confrontati sui seguenti temi: 20 ottobre 2021, *AI: quali rischi per l'autonomia dell'umano*, (Intelligenza + D), con relatori Angelo Alù, Mariagiovanna Gianfreda, Guglielmo Tamburrini, discussant Mario Bochicchio e moderatrice Ilenia Colonna; 27 ottobre 2021, *Immagini del passato, immagini del futuro* (Media + D), con relatori Malvina Giordana, Alma Mileto e Francesco Zucconi, discussant Luca Bandirali e moderatrice Isabella Hernandez; 9 novembre 2021, *Cultural Heritage & Digital Humanities: sfide di accessibilità* (Arte + D), con relatori Eva Degl'Innocenti, Lucio Tommaso De Paolis, Anna Maria Marras, Paola Moscati, discussant Grazia Semeraro e moderatrice Carola Gatto; 10 novembre 2021, *Textual scholarship: forme, strumenti, metodi* (Testo + D), con relatori Marina Buzzoni, Tiziana Mancinelli, Federico Meschini, Andreas Speer, discussant Fabio Ciracì e moderatrice Giulia Miglietta; 12 novembre 2021, *Politiche pubbliche per la costruzione di un ecosistema digitale* (Diritto + D), con relatori Bianca Bronzino, Mino Elefante, Claudia Morini, discussant Marco Mancarella e moderatore Marco Giannotta; 17 novembre 2021, *Tecnologia e umano: quale futuro per la conoscenza* (Filosofia + D), con relatori Simona Chiodo, Riccardo Fedriga, Cristina Marras e Viola Schiaffonati, discussant Fabio Ciracì e moderatrice Patrizia Miggiano; 24 novembre 2021, *Costruire mondi possibili: i videogiochi e le realtà sociali* (Media + D), con relatori Donata Bologna, Marco-Benoît Carbone, Riccardo Fassone e Pietro Luigi Iaia, discussant Luca Bandirali e moderatrice Alessia De Blasi.

La risposta alla *call for papers* è stata, ci pare, all'altezza delle aspettative: sono giunte 86 proposte, con una media di paper accettati del 77%, esattamente 18 paper e 5 poster accettati nell'area Testo+D, 3 paper accettati nell'area Arti+D, 11 paper e 2 poster in area Filosofia+D, 5 paper e 1 poster per Contenuti+D, infine 7 paper per Intelligenza+D, per un totale di 44 paper e 8 poster. Già da una rapida lettura dei titoli si evince non solo la molteplicità dei temi ma anche la varietà degli approcci metodologici, che attestano declinazioni interne anche alle medesime aree tematiche. Infine, per garantire una selezione dei contributi conforme alle aree di ricerca selezionate, abbiamo identificato la figura dei direttori di area, ai quali è stato assegnato il compito di individuare i revisori più adeguati ai temi dei contributi da revisionare, per un'analisi competente e puntuale: per l'area testo, Roberto Rosselli Del Turco; per l'area arti, Grazie Semeraro; per l'area filosofia, Fabio Ciracì; per l'area contenuti, Luca Bandirali e Marco Mancarella; per l'area intelligenza, Mario Bochicchio. A tutti loro va il ringraziamento del Comitato Scientifico e di AIUCD. Una tale suddivisione del lavoro e il supporto della piattaforma digitale *conference* hanno permesso di seguire con efficacia tutto il processo di selezione dei contributi: individuare i revisori idonei, confrontare le valutazioni e richiedere pareri ulteriori in caso di dubbio, controllare che le modifiche richieste agli autori in fase di revisione fossero correttamente apportate alla versione finale del *paper*, selezionare i contributi da presentare alla conferenza. Ciascun contributo è stato valutato da almeno due *referee* in caso di giudizio positivo, almeno tre in caso di giudizio incerto o di giudizi discordanti, o parere negativo. I 75 revisori hanno svolto un lavoro fondamentale di revisione che ha garantito una selezione seria e competente, assicurando al convegno dell'AIUCD la qualità delle proposte e il riconoscimento internazionale duramente conquistato dall'Associazione in questi undici anni di attività.

AIUCD2022 è patrocinato dalla Regione Puglia, dalla Provincia di Lecce e dalla Città di Lecce, la qual cosa è certo indice di una certa sensibilità territoriale ai temi della cultura e della innovazione. Inoltre, il convegno è stato sponsorizzato da: Dipartimento di Informatica dell'Università degli Studi di Bari, AFP – Apulia Film Commission, Il Teatro Pubblico Pugliese, CINI – Consorzio Interuniversitario nazionale per l'Informatica, SFI-Società Filosofica Italiana, il Teatro Pubblico Pugliese e Argo Software, che hanno generosamente sovvenzionato l'iniziativa.

Purtroppo, come nella scorsa edizione, nonostante il ricorso ai vaccini, anche quest'anno il *covid* ha ripreso a correre, improvvisamente, a poco più di una settimana dal Convegno, previsto per il 19-21 gennaio 2022, e ci ha costretti a rinviare il Convegno alla prossima estate. La scelta è stata sofferta e sicuramente ha determinato disagi, ma abbiamo inteso dare priorità alla sicurezza e alla salute pubblica, pur in assenza di decreti restrittivi o limitazioni governative all'attività convegnistica. Non abbiamo inteso proporre invece il convegno in modalità *online*, perché non abbiamo voluto rinunciare al nostro amato convegno in presenza. La virtuosa trasposizione in modalità digitale di AIUCD2021 offerta, in emergenza, per il Convegno di Pisa è stata sicuramente un esperimento riuscito. Tuttavia, dopo due anni di pandemia, il Comitato Scientifico, di concerto con il Direttivo AIUCD, ha reputato opportuno scegliere comunque di rinviare, per privilegiare il convegno in presenza, senza ovviamente rinunciare ai vantaggi offerti dalla modalità ibrida. Un ulteriore convegno solo in remoto avrebbe altrimenti gravato immancabilmente sugli aspetti sociali e relazionali, per nulla secondari, che costituiscono la vera sostanza del convegno nazionale, rendendolo un luogo di confronto vivo, un'insostituibile occasione di relazione e di partecipazione attiva. Siamo dell'opinione che il digitale debba rappresentare un'opportunità, non già una dimensione sostitutiva ed esclusiva, ma complementare e inclusiva.

Il Convegno previsto per il 19-21 gennaio 2022 indicava la partecipazione di prestigiosi studiosi che arricchivano la proposta tematica di AIUCD2022, che intendiamo confermare anche per il rinvio di giugno. Innanzitutto, i nostri *keynote*: Luciano Floridi – Professore Ordinario di filosofia ed etica dell'informazione presso l'Oxford Internet Institute e direttore del Digital Ethics Lab dell'Università di Oxford, nonché Professore di Sociologia della comunicazione presso l'Università di Bologna – inaugurerà il convegno con una lezione su *Semantic capital: its nature, value, and preservation*; Maurizio Ferraris – Professore Ordinario di filosofia teoretica presso la Facoltà di Lettere e Filosofia dell'Università degli Studi di Torino e noto studioso della *documedialità* – concluderà i lavori con una lezione intitolata *Webfare*. Si aggiungeranno gli *invited speaker* che, per ogni giorno della conferenza, sviluppano un tema specifico del convegno: Maria Grazia Mattei – umanista, critica d'arte e direttrice di *Meet the Media Guru* – si soffermerà sull'*Arte digitale: storia e panoramica attuale*; Gino Roncaglia – Professore Associato dell'Università Roma Tre, esperto di digitale e cultura del libro, consulente RAI – discuterà di *Simulismi;* Anna Bisogno – Professore Associato di Cinema Radio e Televisione dell'Università Telematica Mercatorum – analizzerà *La rete-visione. Televisione e schermi nell'era digitale;*

infine, Riccardo Fedriga – Professore Associato dell'Università di Bologna, esperto di editoria digitale, storico delle idee – esaminerà le *Fruttuose debolezze. Fragilità e indeterminismi digitali.*

Lavoreremo affinché il programma, così faticosamente costruito per gennaio, non subisca variazioni strutturali. Inoltre, al posto del consueto *Book of Abstracts,* per l'edizione del 2022 l'AIUCD ha scelto di pubblicare i *Proceedings*, come segno tangibile di un processo di aggiornamento continuo del Convegno Nazionale e di crescita intellettuale dell'Associazione. Essi vedono la luce nonostante il rinvio del convegno in presenza, per fornire una base alla discussione che si svolgerà questa estate, con la consapevolezza che gli studi pubblicati fotografano lo stato dell'arte, ma che la ricerca è in continua evoluzione. Quindi, in sede di convegno, faremo i conti con i progressi avvenuti nei mesi trascorsi dalla pubblicazione dei *Proceedings*, di cui terremo conto per l'eventuale pubblicazione dei *selected papers*.

Vorremmo chiudere la prefazione rivolgendo un particolare ringraziamento ai membri del Comitato Scientifico e, *last but not least*, esprimendo profonda gratitudine ai componenti del Comitato di programma, coordinati da Federica Epifani: tutte giovani e promettenti energie intellettuali a cui è dedicato il presente volume di *Proceedings*, non a caso edito a cura di Giulia Miglietta e Carola Gatto.

Fabio Ciracì
Mario Bochicchio

Conservazione e fruizione di banche dati letterarie: l'archivio della poesia italiana dell'Otto/Novecento di Giuseppe Savoca

Antonio Di Silvestro[1], Christian D'Agata[2], Giuseppe Palazzolo[3], Pietro Sichera[4]

[1]Università degli Studi di Catania - silvanto [at] unict.it

[2]Università degli Studi di Catania - christian.dagata [at] studium.unict.it

[3]Università degli Studi di Catania – giuseppe.palazzolo [at] unict.it

[4]CNR-ISTC – pietro.sichera [at] istc.cnr.it

ABSTRACT

In un periodo nel quale si va sempre più verso l'interoperabilità delle banche dati, in particolare di quelle letterarie, si pone il problema di ovviare al rischio di obsolescenza degli archivi testuali dei decenni precedenti, assicurando non solo la corretta migrazione dei dati nelle nuove piattaforme ospitanti, ma anche la loro persistenza digitale. È una questione che interessa l'Archivio della poesia italiana dell'Otto/Novecento, ideato negli anni '80 da Giuseppe Savoca presso l'Università di Catania, a cui l'interazione tra gli elementi lemmatizzati e la loro codifica in XML-TEI offrirà continuità e nuove potenzialità di consultazione e interrogazione.

PAROLE CHIAVE

FAIR principles, lemmatizzazione concordanziale, vocabolario poetico italiano, data retrieval, archivi digitali.

INTERVENTO

1. INTRODUZIONE

In *Se una notte d'inverno un viaggiatore* ([4]) Lotaria dà una delle prime descrizioni letterarie dell'applicazione di processi elettronici alle pratiche di critica del testo. «La lettura elettronica mi fornisce una lista delle frequenze, che mi basta scorrere per farmi un'idea dei problemi che il libro propone al mio studio critico». Tra i modelli di letture e di misletture che l'autore si diverte a censire, quello presentato da Lotaria è la parodia dell'uso delle concordanze nella critica letteraria. Il romanzo di Calvino viene pubblicato nel 1979 e cita espressamente gli *Spogli elettronici dell'italiano letterario contemporaneo* ([2]). Negli stessi anni, a Catania, Giuseppe Savoca elabora programmi e metodi informatici di spoglio lessicale, con il supporto di un software proprietario. Lo studioso riesce a delegare a un sistema IBM/AS400 un'operazione altamente artigianale quale la realizzazione di una concordanza lemmatizzata, in un periodo nel quale, con l'eccezione di alcune pregiate concordanze cartacee (ad es. quella del *Canzoniere* petrarchesco realizzata dall'Accademia della Crusca ([11])) e dell'opera fondamentale di Padre Busa ([3]) a circolare erano soprattutto formari bruti, con pochissimi dati o inferenze di tipo statistico, di non secondaria importanza nell'interpretazione dei testi letterari.

Il processo computazionale che sta alla base delle concordanze lemmatizzate prodotte da Savoca e dalla sua équipe di ricerca ha prodotto finora degli output cartacei dotati di esaustività (le concordanze comprendono infatti anche i lemmi sinsemantici – articoli, preposizioni, congiunzioni) e di strumenti di studio statistico assai ricchi e diversificati[1].

L'esito più prestigioso di questa operazione editoriale è confluito nel *Vocabolario della poesia italiana del Novecento* [8]. Tuttavia, le stampe cartacee non riuscivano a tenere il passo con l'incremento quantitativo e qualitativo dell'archivio, dove intanto confluivano anche i poeti 'minori', nonché i corpora costituiti dalle traduzioni poetiche. Inoltre, l'ingresso di autori del Sette/Ottocento (Parini, Foscolo, Manzoni, il Leopardi 'puerile') e di una nuova concordanza del *Canzoniere* petrarchesco ([10]) rendevano l'archivio elettronico un paradigma esemplare della storia della poesia italiana da Petrarca al Novecento. Un simile patrimonio necessita di essere preservato e messo a disposizione della comunità scientifica nazionale e internazionale. Esigenze di calcolo, modalità di visualizzazione innovative e aggiornamento delle metodologie sia informatiche quanto linguistiche consentono, e in un certo senso obbligano, la migrazione di questi dati verso sistemi diversi, adottando anche un punto di vista computazionale moderno e portando il progetto verso la riusabilità e l'interoperabilità. È il nucleo dei principi FAIR ([13]): l'obiettivo non è più la semplice gestione corretta dei dati (da un

[1] Tutte le più significative concordanze dei poeti dell'Otto/Novecento (da Leopardi a Montale, da D'Annunzio a Pavese) sono state ospitate nella prestigiosa collana "Strumenti di Lessicografia Letteraria Italiana" dell'editore Olschki, che ha finora stampato quasi trenta volumi di concordanze e studi teorici ad esse collegati. Si ricordano in particolare la prima su Gozzano ([7]) e il volume teorico e metodologico *Lessicografia letteraria e metodo concordanziale* ([9]).

punto di vista tanto teorico quanto pratico, informatico quanto lessicografico), ma l'integrazione e il riutilizzo della conoscenza dopo la pubblicazione dei dati stessi. In tale contesto *Findability*, *Accessibility*, *Interoperability*, e *Reusability* diventano aspetti cardine nella gestione del processo di migrazione, la cui analisi e le cui ricadute scientifiche e tecnologiche sono oggetto di questo articolo.

2. LA CONCORDANZA NELL'ARCHIVIO DELLA POESIA ITALIANA DI G. SAVOCA: PREMESSE TEORICHE E FUNZIONAMENTO

Negli anni '80 i programmi esistenti in commercio si basavano essenzialmente, se non esclusivamente, sulla semplice messa in ordine alfabetico delle forme brute di un testo[2]; non erano in grado di eseguire una corretta lemmatizzazione, in quanto non distinguevano le forme omografe, né, tanto meno, prevedevano la classificazione grammaticale dei lemmi o la generazione automatica di indici statistici. Le concordanze realizzate secondo il modello Savoca, in quanto finalizzate alla lettura paradigmatica di un testo e/o allo studio del sistema linguistico di un autore, forniscono elementi oggettivi, formalizzati e quantificati, relativi a tutto il lessico, comprese le parole cosiddette funzionali – articoli, congiunzioni, preposizioni –, che occupano circa il 30% della lingua di qualunque testo.
I programmi di lemmatizzazione ideati da Savoca lavorano sulla piattaforma IBM AS/400[3], uno dei pochi sistemi che negli anni '80 poteva essere in grado di elaborare la quantità di informazioni necessaria al lavoro di catalogazione. In *Figura 1* viene mostrata il menu di scelta programmi del sistema Savoca su AS/400.

Figura 1. Menù principale del software concordanziale su AS400

Secondo lo schema di Savoca, il programma importa il testo da analizzare attribuendo ad ogni sua riga un codice univoco a dimensione fissa. Prendendo come esempio la riga del "File versi" (riprodotto in *Figura 2*) "I corpi di Clitennestra e di Egisto", si vede come alla riga vengono attribuiti un'etichetta "V" (indicante il tipo di riga, verso in questo caso), e un codice composto da "PA" (due lettere indicanti il codice dell'autore, Pasolini), "AA" (codice attribuito all'opera, *Pilade*) e un numero di otto cifre (la posizione del verso all'interno dell'opera).

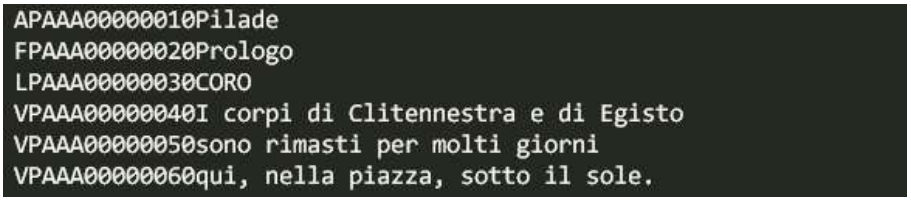
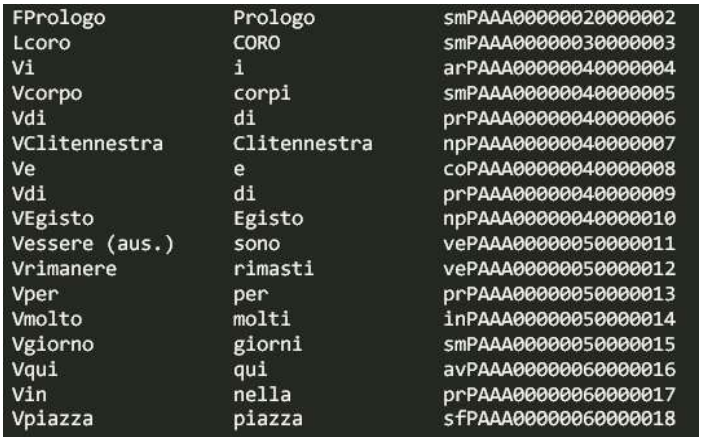


Figura 2. Esempio di "File versi" su AS400

Grazie a questa struttura tutte le righe da concordare risultano univocamente determinate.
A partire dal "File versi" viene generato un "File lemmi", in ordine sequenziale o alfabetico. In *Figura 3* è possibile vedere il "File lemmi" relativo al "File versi" di *Figura 2*. Anche qui ogni riga risulta strutturata. Prendendo come esempio la riga relativa all'occorrenza "Clitennestra", si vede come tale riga è composta da "V" (il codice indicante il tipo della riga nel "File versi"), "Clitennestra" (il lemma, l'entrata lessicografica da cui deriva l'occorrenza), "Clitennestra" (l'occorrenza,

[2] Si vedano a tal proposito, a titolo esemplificativo, gli output generati dal software "LIZ 3.0 - A cura di Pasquale Stoppelli ed Eugenio Picchi - 1998"
[3] Con IBM AS/400 (Application Server/400) si intende quella famiglia di computer *midrange* progettati da IBM alla fine del secolo scorso che consentissero di ottenere alte prestazioni di calcolo a prezzi relativamente contenuti.

cioè la parola come è scritta nel testo), “np” (il codice della categoria grammaticale del lemma, nome proprio in questo caso), “PAAA00000040” (codice della riga come presente nel “File versi”), “000007” (contatore progressivo delle occorrenze nel file). Si noti quindi come sia il “File versi” che il “File lemmi” siano posizionali, impongano cioè l’uso di un certo numero di caratteri per ogni campo.



FPrologo	Prologo	smPAAA00000020000002
Lcoro	CORO	smPAAA00000030000003
Vi	i	arPAAA00000040000004
Vcorpo	corpi	smPAAA00000040000005
Vdi	di	prPAAA00000040000006
VClitennestra	Clitennestra	npPAAA00000040000007
Ve	e	coPAAA00000040000008
Vdi	di	prPAAA00000040000009
VEgisto	Egisto	npPAAA00000040000010
Vessere (aus.)	sono	vePAAA00000050000011
Vrimanere	rimasti	vePAAA00000050000012
Vper	per	prPAAA00000050000013
Vmolto	molti	inPAAA00000050000014
Vgiorno	giorni	smPAAA00000050000015
Vqui	qui	avPAAA00000060000016
Vin	nella	prPAAA00000060000017
Vpiazza	piazza	sfPAAA00000060000018

Figura 3. Esempio di “File lemmi” su AS/400

Il software dispone di un dizionario di macchina composto da circa 8.000 forme che gli consentono di avere una sensibilità del 70%, una precisione media del 90% e quindi un punteggio F1 di 0,78[4]. Si rende dunque necessario l’intervento umano per correggere e completare il processo concordanziale, operazione sicuramente non banale. Quantunque infatti la prima regola lessicografica del concordatore sia quella di adeguarsi al modello adottato dai dizionari (il verbo all’infinito, i sostantivi al singolare, il superlativo degli aggettivi ricondotti al grado positivo, ecc.), egli, tuttavia, si trova a dover elaborare un dizionario del tutto speciale perché limitato a un’opera (o a un corpus), con particolarità linguistiche, testuali e grafiche che non sempre possono trovare riscontro nell’uso medio attestato da un dizionario. Per di più, in un’opera poetica può esserci tutto un corredo di notizie, riferimenti, nomi propri, titoli, citazioni, ecc., che non rientrano fra i lemmi accolti nei dizionari, e che perciò pongono al concordatore problemi di lemmatizzazione a volte molto complessi.
Portata a termine la fase non automatizzata del processo, dall'unione del “File versi” e del “File lemmi” viene generata la concordanza vera propria.
Oltre alla realizzazione della concordanza, l'AS/400 consente la produzione di altri utili strumenti lessicografici di consultazione, come liste di frequenza (lemmi o forme in ordine alfabetico, lemmi o forme per numero di occorrenze in ordine crescente o decrescente, lemmi ordinati per categoria grammaticale, retrograde di forme e/o lemmi, utili ai fini dello studio delle rime), liste di lemmi specifici, liste di lemmi comuni a più opere e a più autori, quadri statistici (totale dei lemmi, delle forme e delle occorrenze del testo concordato, frequenza assoluta e frequenza relativa delle categorie grammaticali, ecc.).
Il numero di concordanze create con questo sistema dagli anni ’80 a oggi è notevole. L'archivio contiene oltre 1.400 file tra “File versi”, “File lemmi” e concordanze generate, per un totale di diversi milioni di record.
Si evince da questa descrizione che l’archivio dati descritto risulta di tipo *flat-file*, in cui la formattazione dei dati risulta essenziale, limitando però notevolmente l’interoperabilità e la riusabilità degli stessi. Inoltre la presenza sia nel “File versi” che nel “File lemmi” del campo “Tipo verso” contravviene alle regole sulla normalizzazione ([1]) dei database[5]. Diventa quindi essenziale ripensare questo sistema concordanziale sia come progettazione che come modalità di fruizione.

[4] Con *precisione* e *sensibilità* (*precision* e *recall*) si intendono due indicatori per valutare la qualità di un modello decisionale. Nel modello preso in esame, la precisione indica il rapporto tra i casi corretti e quelli che il modello ha riconosciuto, la sensibilità indica il rapporto tra i casi corretti rispetto all’intero *corpus*. La media armonica di questi due indicatori (2*((P*S)/(P+S))) viene chiamata punteggio F1 ed è un metro ampiamente utilizzato per valutare complessivamente i sistemi di apprendimento automatico.
[5] Nella teoria dei database le forme normali sono delle regole teoriche che garantiscono la corretta progettazione di una base di dati relazionale, evitando episodi di ridondanza e incoerenza.

3. LA CONCORDANZA SU WINDOWS, MACOSX, WEB: LIOTROCONCORD

Da alcuni anni il CINUM[6] ha implementato un nuovo software per la generazione delle concordanze, chiamato LiotroConcord (versione 2). Il software, sviluppato a partire dal modello concordanziale di Savoca, è un database relazionale puro, creato utilizzando la piattaforma 4D.

4D[7] è una piattaforma RAD[8] per DBMS[9] che integra al suo interno un motore per database SQL[10], un linguaggio di programmazione proprietario e un server web: un modello LAMP[11] riunito in un unico strumento.

A queste caratteristiche se ne aggiungono varie altre come il supporto nativo SVG[12], i comandi DOM e SAX per l'analisi e la scrittura di file XML[13], un interprete PHP[14].

La creazione di un nuovo software si è resa necessaria per vari motivi. Ad esempio le esperienze maturate permettono di avere un progetto tecnicamente più avanzato, l'utilizzo di personal computer Windows o MacOSX permette di creare interfacce utente moderne e *user-friendly*, facilitando l'usabilità e l'interoperabilità con altri dati e altri software.

LiotroConcord è stato progettato seguendo la teoria delle forme normali ([12]), in modo da garantire correttezza di sviluppo e facilità di mantenimento. Il software è utilizzabile indifferentemente e contemporaneamente sia su sistemi Windows che MacOS, nonché via browser web.

Al centro di LiotroConcord viene posta la parola nelle sue varie declinazioni: occorrenza, forma e lemma. Tutte le concordanze generate dal sistema sono strettamente connesse, indipendentemente dall'autore o dal genere (prosa, poesia, teatro). La generazione della concordanza viene resa quanto più possibile automatica passando attraverso una metodologia ormai consolidata. L'utente, attraverso un editor di testi esterno, compone il corpus da importare nel sistema. Il file così generato (.txt o .docx) viene sottoposto da LiotroConcord ad un'analisi preventiva del file secondo la codifica UTF-8[15]. Preparato il testo, LiotroConcord lo analizza, lemmatizzando correttamente la quasi totalità del testo, grazie ad un dizionario di macchina di quasi 300.000 forme, disambiguando inoltre le forme omografe[16]. Il software utilizza un sistema di *machine learning* ad apprendimento con rinforzo: la correzione manuale della concordanza realizzata automaticamente fornisce dei feedback che gli consentono di imparare a correggere i propri errori. Il sistema di riconoscimento utilizza varie tecniche, come ad esempio la valutazione degli "intorni" (l'occorrenza o i gruppi di occorrenze che precedono/seguono la forma omografa) o la presenza di imitazioni, utilizzo cioè della stessa sequenza di occorrenze/lemmi/categorie grammaticali nelle opere dello stesso autore o nell'intero database. LiotroConcord acquisisce ed impara: attualmente presenta punte negli indici di precisione e sensibilità del 98% con un indice F1 che si attesta sempre a 0,97 (valori nettamente più elevati rispetto al sistema su AS/400).

Per il controllo e la correzione della concordanza, LiotroConcord offre un editor a interfaccia grafica appositamente progettato per rendere veloce ed accurato il processo di controllo, anche quando questo viene svolto contemporaneamente da più utenti.

---

[6] Il CINUM (Centro di Informatica Umanistica) è un centro interdipartimentale istituito presso l'Università di Catania nel 2017. Riunisce studiosi di vari ambiti disciplinari (dalla linguistica alla filosofia del linguaggio, dalla filologia alla sociologia della comunicazione), e attualmente è impegnato, oltre che nell'Edizione Digitale dell'Opera Omnia di Luigi Pirandello (www.pirandellonazionale.it), in diversi progetti di ricerca su fondi di Ateneo riguardanti sia la messa in rete dell'archivio della poesia italiana dell'Otto/Novecento, realizzato da Giuseppe Savoca e dalla sua equipe di ricerca a partire dagli anni '80, sia la realizzazione di edizioni-archivio interrogabili della narrativa e poesia italiana del Novecento.

[7] La piattaforma 4D è sviluppata e mantenuta dall'omonima società francese 4D SAS. Per un elenco di caratteristiche si veda https://us.4d.com/new-4d-v19lts.

[8] RAD (Rapid Application Development) è una metodologia di sviluppo informatico basata sulla costruzione di prototipi, modelli di sviluppo iterativi e utilizzo di interfacce grafiche.

[9] Con DBMS (Database Management System) si intende un software per la realizzazione di database. Infatti, alla realizzazione e interrogazione della base dati, aggiunge altre caratteristiche, come linguaggi di programmazione e gestori dell'interfaccia.

[10] SQL (Structured Query Language) è un linguaggio per database inizialmente sviluppato da Donald D. Chamberlin e Raymond F. Boyce ([5]) basato sul modello relazionale introdotto da Edgar F. Codd.

[11] LAMP è l'acronimo di 'Linux, Apache, MySQL, PHP', uno dei tipici ambienti per lo sviluppo di applicazioni web, dove Linux è il sistema operativo, Apache è il software che funge da web server, MySQL è il motore del database e PHP il linguaggio di programmazione.

[12] SVG (Scalable Vector Graphics) è un formato che supporta oggetti di grafica vettoriale.

[13] XML (eXtensible Markup Language) è un linguaggio di markup che consente la codifica di documenti in un formato leggibile sia dall'uomo che da una macchina. I documenti XML sono strutturati e DOM (Document Object Model) e SAX (Simple API for Xml) sono linguaggi usati per accedere a tale struttura.

[14] PHP (PHP Hypertext preprocessor) è un linguaggio di scripting per pagine web dinamiche.

[15] UTF-8 è un sistema di codifica per i caratteri Unicode. Mentre ASCII permette di codificare solo 128 caratteri e le codifiche ISO-8859 alcuni set di caratteri stampabili, UTF-8 amplia le possibilità di codifica facendo uso di codici fino a 4 byte. Lo standard è reperibile al link https://datatracker.ietf.org/doc/html/rfc3629

[16] Forme omografe come ad esempio 'sale' (sostantivo maschile singolare, sostantivo femminile plurale o voce verbale) oppure 'che' (congiunzione o pronome relativo).

Il sistema è in grado di effettuare le analisi linguistiche e generare gli output già presenti nel sistema Savoca (concordanza, liste di frequenza, ecc.), nonché altre analisi come l'iperconcordanza di più corpora, l'intersezione lessicale (per lemma e per forma tra parti dello stesso corpus o tra corpora differenti), la valutazione delle co-occorrenze.
È stato sviluppato anche un sistema di interrogazione tramite pagine web[17] attraverso cui è stato possibile pubblicare testi lemmatizzati prodotti dall'équipe di ricerca di Giuseppe Savoca e successivamente dall'équipe del CINUM. Essendo 4D, come detto in precedenza, un ambiente integrato, non è necessario trasferire i dati dall'ambiente di concordanza ad interfaccia utente all'ambiente web: i dati risultano sempre già pronti per essere pubblicati sul web e ogni modifica apportata attraverso l'interfaccia utente è immediatamente visibile nel sistema di interrogazione.

4. PROCESSO DI TRASFORMAZIONE

La migrazione delle concordanze generate su AS/400 verso il nuovo software LiotroConcord è stata svolta analizzando e risolvendo varie problematiche. Innanzitutto la codifica dei caratteri dei vecchi file risultava non standard. Per questo motivo si è dovuto ricorrere ad un'analisi puntuale dei file esportati dall'AS/400 correggendo le difformità che si evidenziavano, soprattutto per quanto riguardava i caratteri con codice ASCII superiore a 128 (le lettere accentate).
I "File lemmi" e i "File versi" necessitavano di una conversione dalla codifica proprietaria del sistema AS/400 a quella UTF-8. Per effettuare questo passo, si è utilizzato il vocabolario di macchina di LiotroConcord che, analizzando le parole sconosciute (ad esempio "perch?" al posto di "perché"), è riuscito ad effettuare in automatico l'aggiornamento della codifica.
Codificati i due file in maniera corretta, si è proceduto alla loro importazione combinata. Il "File versi" è stato importato tenendo traccia del codice attribuito al verso dal sistema AS/400. Di seguito sono state cercate all'interno del "File lemmi" le forme presenti in un certo verso; attraverso un algoritmo di *pattern matching*[18] sono state confrontate le possibili occorrenze del verso importato con le occorrenze che il "File lemmi" attribuiva a quel verso. Il processo si è rivelato corretto in maniera automatica al 99,9%[19].
È stato così possibile non solo attribuire le forme ai versi relativi, ma anche riscontrarne l'effettiva posizione, in modo da poter utilizzare le funzioni di LiotroConcord che analizzano le co-occorrenze. Il processo si è rivelato esaustivo, essendo infatti in grado di trasferire integralmente le concordanze dal vecchio al nuovo sistema.

5. RICADUTE SCIENTIFICHE. UN MODELLO PER IL FUTURO

L'obiettivo principale del processo di trasformazione è stato quello di incanalare le concordanze prodotte da Savoca all'interno dei principi FAIR: evitare l'obsolescenza digitale delle concordanze (dati reperibili e riutilizzabli) e garantirne l'accessibilità e l'interoperabilità. A tale scopo è attivo presso il CINUM il progetto ALPION[20], che si pone come obiettivo quello della trasmigrazione dei dati in un sistema interoperabile, permettendo un'interrogazione del corpus da parte degli studiosi. Il problema di fondo è tuttora al centro del dibattito nella comunità scientifica, ma certamente un modo per superarlo è quello di condividere i contenuti in formati aperti come XML-TEI[21]. La concordanza diventa così un prodotto che, come nella menzionata Edizione Digitale dell'Opera omnia di Pirandello[22], è liberamente consultabile sia attraverso un form d'interrogazione sul web, sia attraverso un pdf scaricabile ([6]). La *Figura 4* mostra uno *specimen* dell'output in XML-TEI generato dalla ricerca del lemma «abate» all'interno dell'*Enrico IV* di Luigi Pirandello:

[17] Il sistema di interrogazione è stato implementato per l'Edizione Nazionale dell'Opera omnia di Luigi Pirandello (https://vocabolari.pirandellonazionale.it/) e per il progetto ALPION (https://alpion.unict.it/vocabolario/ricerca/).
[18] È stata implementata una variante del metodo ad indici che, basandosi sul *preprocessing* del testo, risulta molto performante. Dati *m* la lunghezza del pattern e *Z* il numero di occorrenze di un certo pattern, l'algoritmo impiega O(m+Z) per trovare tutte le occorrenze.
[19] Si è scelto volutamente di marcare come "Da controllare" le occorrenze omografe all'interno di un verso, quando tali occorrenze provenivano da lemmi differenti.
[20] https://alpion.unict.it/
[21] TEI Consortium. *TEI P5: Guidelines for Electronic Text Encoding and Interchange.* Version 4.3.0. Last updated on 31st August 2021. TEI Consortium. http://www.tei-c.org/Guidelines/P5/.
[22] https://www.pirandellonazionale.it/

Figura 4. Risultato della ricerca del lemma «abate» all'interno dell'Enrico IV di Luigi Pirandello

In questo modo la concordanza diventa facilmente importabile e riutilizzabile in qualunque altro sistema per basi di dati. Avendo portato a termine la realizzazione del sistema di migrazione dei dati dal sistema AS/400 ed essendo già in fase avanzata l'importazione dei dati stessi verso il software LiotroConcord, uno degli obiettivi che ALPION si pone adesso è quindi quello di permettere il libero accesso agli studiosi attraverso un sistema di interrogazione dei dati su interfaccia web e l'esportazione dei risultati in file XML-TEI (elaborati direttamente da LiotroConcord) che rispettino le informazioni strutturate della concordanza.
Un ulteriore possibile output è quello della lemmatizzazione integrale del testo, laddove – è bene chiarirlo – siano già scaduti i diritti d'autore e sia possibile rendere disponibile alla comunità scientifica il testo per intero. Il file così condiviso sarebbe quello della lemmatizzazione, rendendola in tal modo integralmente interoperabile. A mo' di esempio mostriamo in *Figura 5* il possibile output di un verso de *Le ceneri di Gramsci* di Pier Paolo Pasolini.

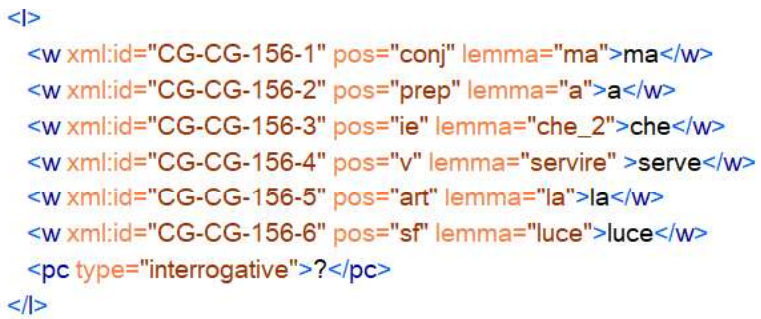


Figura 5. Un possibile output della concordanza de Le ceneri di Gramsci di Pier Paolo Pasolini in XML-TEI

In questo caso a essere reso disponibile non sarà tanto il prodotto finale, ma la lemmatizzazione *in fieri*, che può essere modificata, integrata, revisionata da qualunque studioso.

BIBLIOGRAFIA

[1] Albano, Antonio, Giorgio Ghelli, e Renzo Orsini. 1997. *Basi di dati relazionali e a oggetti*. Bologna: Zanichelli.
[2] Alinei, Mario. 1978. *Spogli elettronici dell'italiano letterario contemporaneo*. 3 voll. Bologna: Il Mulino.
[3] Busa, Roberto. 1974. *Index Thomisticus Sancti Thomae Aquinatis Operum Omnium Indices et concordantiae*. Vol. 56. Stuttgart: Frommann-Holzboog.
[4] Calvino, Italo. 1994. *Se una notte d'inverno un viaggiatore*. Vol. 2. Romanzi e racconti. Milano: Mondadori.
[5] Chamberlin, Donald D., e Raymond F. Boyce. 1974. «SEQUEL: A Structured English Query Language». In *ACM SIGFIDET*.
[6] Giuffrida, Milena, Christian D'Agata, Laura Giurdanella, e Pietro Sichera. 2021. «Pirandello Nazionale: per un nuovo modello di edizione digitale, collaborativa e integrata». In *Boschetti, Federico; Del Grosso, Angelo Mario; Salvatori, Enrica; (edd.); AIUCD 2021 - DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale. Raccolta degli abstract estesi della 10a conferenza nazionale*. Pisa.
[7] Savoca, Giuseppe. 1984. *Concordanza di tutte le poesie di Guido Gozzano*. Firenze: Olschki.
[8] ———. 1995. *Vocabolario della poesia italiana del Novecento*. Bologna: Zanichelli.
[9] ———. 2000. *Lessicografia letteraria e metodo concordanziale*. Firenze: Olschki.

[10] Savoca, Giuseppe, e Bartolo Calderone. 2011. *Concordanza del Canzoniere di Francesco Petrarca*. 2 voll. Firenze: Olschki.
[11] Ufficio lessicografico dell'Opera del Vocabolario. 1971. *Concordanza del Canzoniere di Francesco Petrarca*. 2 voll. Firenze: Firenze: Accademia della crusca.
[12] Ullman, Jeffrey D. 1988. *Principles of Database and Knowledge - Base System*. USA: Computer Science Press Inc.
[13] Wilkinson, Mark D., Michel Dumontier, et al. 2016. «The FAIR Guiding Principles for scientific data management and stewardship». *Sci Data 3* 160018.