

Historical Network Analysis & Htr Tool

Per un approccio storico metodologico digitale all'archivio Biscari di Catania

Salvatore Spina

Università degli Studi di Catania, Italia
salvatore.spina@unict.it

Abstract

L'archivio della famiglia Paternò Castello, principi Biscari, rappresenta un patrimonio documentario la cui struttura e complessità ha attirato studiosi diversi, guidati dalla necessità di comprendere la “capitale” etnea, la sua storia, i suoi attori e come questi determinarono la sua struttura culturale. L'importante storiografia, però, non è mai riuscita a rispondere esaustivamente a tutte le domande. Oggi, la prospettiva metodologica digitale e l'uso che essa pone all'attenzione del *digital historian*, di specifiche tecnologie informatiche, quali l'Handwritten Text Recognition (Transkribus), l'Historical Network Analysis, o web-based tools, come Keyphrase-Digger, apre nuovi spazi interpretativi. L'upgrade delle fonti archivistiche alla loro versione digitale potrà consentire quello studio “quantitativo-qualitativo” che porterebbe nuova luce su soggetti ed eventi, sulle reti e le risorse, sulle anime e le decisioni politiche che scrissero la storia di Catania e delle Sicilie.

The Paternò Castello (Princes of Biscari) Archive – amongst one of the most crucial family archives –, is the key – historical heritage to write the history of Catania, of the parties concerned and how the latter determined the cultural structure of the city. Most scholars, however, have never managed to exhaust all the questions and answers. Today, the digital methodological perspective and its ITC tools, such as Handwritten Text Recognition (Transkribus), Historical Network Analysis, and web-based tools, such as Keyphrase-Digger, open up new historical paths. Upgrading archival sources to their digital version will allow “quantitative-qualitative” research that would shed new light on historical actors and events, networks and resources of the history of Catania and the Two Sicilies.

Parole chiave: Transkribus, Filemaker, Gephi, Principe Ignazio, Principessa Anna, Michele Maria Paternò.

Keywords: Transkribus, Filemaker, Gephi, Prince Ignatius, Princess Anne, Michele Maria Paternò.

1. Premessa

Se il processo della digitalizzazione e informatizzazione rappresenta la risposta alla necessità di un percorso che rivaluti gli statuti disciplinari, la domanda su come tale rivoluzione stia rimodulando il mestiere di storico resta ancora da definirsi in maniera dettagliata. Certamente, però, il discorso epistemologico ha risolto molte criticità e ha portato i significati di questa disciplina dentro la dimensione della *biotecnocenos*¹ ([51]).

La rivoluzione tecnoinformatica induce la metodologia della Storia ad una riflessione sui suoi paradigmi, con la finalità di portare gli storici dentro gli spazi del dibattito sul futuro digitale delle Scienze, soprattutto alla luce dello sviluppo di metodi e tecnologie e intelligenze artificiali che consentono di guardare al Passato da un punto di vista innovativo, che si traduce in originali prospettive di ricerca, quindi porre nuove domande al patrimonio documentario. Se, infatti, il mondo analogico-tipografico è caratterizzato dal *close reading*, dall'intimità razionale e riflessiva esistente tra lo storico e le sue fonti, gli approcci digitali ampliano il quadro, permettendo allo studioso di superare molti limiti, soprattutto in termini di quantità di documentazione accessibile, organizzazione delle fonti in database, loro analisi e controllo.

Conseguenza reale della «hyperlearning revolution», così descritta da Lewis Perelman in *School's out* ([47]), è la vastità delle carte con cui lo storico si rapporterà. Sono i «Big Data della Storia», affermazione concreta di Raffaele Santoro ([49]), che porteranno, nell'odierno «domani», a cogliere aspetti invisibili degli eventi del passato.

Certamente, affermazioni ed etichette di tal portata, comportano una presa di posizione: non ricadere nella ridondanza della mera quantificazione. Ma se vera l'idea che il divario tra metodologia quantitativa e qualitativa è stato in parte sintetizzato ([30]), così come dimostrato dalle proposte avanzate nel 2008 dall'Economic and Social Research Council, in collaborazione con l'Higher Education Funding Council, di rilanciare e promuovere lo studio e l'insegnamento degli approcci quantitativi nelle università inglesi, è ancor più vero che gli approcci informatici alla ricerca storica sono in divenire e mostrano le difficoltà derivanti da una formazione quasi assente, in tale direzione, e di un sistema di conservazione delle fonti che non riesce a tradurre in digitale il patrimonio che possiede.

Questa condizione comporta, così, una presa di coscienza che ci obbliga a guardare alle attuali progettualità sul patrimonio storico e culturale, come espressione di quella che io definisco «lunga fase del *digital data storing*», che ha avuto inizio nel momento in cui il Web ha consentito a tutti i suoi utenti di immettere nella rete delle informazioni che, al di là delle validazioni, consentiranno di approfondire quanto narrato, storicamente, fino ad oggi. *Ma ancora non oggi*. Quella della *processabilità* è, a mio avviso, una fase che necessita ancora una puntuale definizione della strutturazione di una «cohesive collection».

Ma, il dado è tratto. Per Nawrotzki e Dougherty, «the digital revolution transformed how we write about the past ([14])», e consentirà, secondo Le Roy Ladurie, «una vera rivoluzione storiografica» ([34]).

Quello dello storico, ovviamente, resta lo scopo di un ragionamento sulla base di quegli «avanzi» (*Überreste*), di quei monumenti (*Denkmäler*) ([15]) e fonti (*Überlieferung*), che sono espressione della ragione umana e della sua relazione con il mondo; e su di essi si ergono il concetto di

1 Il concetto è mio

“testimonianza” e quello di “mestiere di storico”. Il ricorso ai calcolatori, però, invita ad una riconcettualizzazione del significato di «fonte», per tutelarne l'essenza nel momento in cui essa viene codificata per essere analizzata da *tools* informatici.

E se, come afferma Gerard Labuda, la fonte è un «resto psicofisico e sociale ... che rende possibile la ricostruzione scientifica dello sviluppo della società, in tutti i suoi aspetti ([33])», la tecnologia informatica diventa “testimonianza”, ma anche unico mezzo per comprenderla. È l'esponenziale possibilità di calcolo, infatti, che rende la Macchina di Turing lo strumento necessario all'organizzazione e analisi delle informazioni che, al contrario, in un archivio analogico, restano prive di significato.

Per questo motivo, proprio la metodologia della ricerca storica, in ambiente digitale, spinge lo storico all'implementazione di archivi digitali e database ([17];[24];[35]) in grado di rispondere con più efficacia, e con «immenso progresso pratico per via del risparmio di tempo» ([18]), ai bisogni della ricerca, ossia alla necessità di costruire uno spazio semantico dove organizzare tutte quelle fonti che devono rispondere alle domande sulla Storia, creando quel significato storico che le interconnette, consentire una rapida possibilità di ricerca tra i dati, e fornire strumenti per esportare i risultati delle *query* in formati che possono essere successivamente elaborati da algoritmi e IA.

E questo è il percorso metodologico seguito per ricondurre l'attenzione su una casata nobiliare, e su alcuni suoi membri, che mostrano rilevanti profili culturali, che li distinguono dalla “serialità” di chi opera nella loro stessa dimensione sociale e politica, o da chi guarda alla cultura come lato necessario all'ostentazione della propria levatura. L'approccio, per questo motivo, vuole far proprie le categorie di “radicamento”, “costruzione del potere”, “egemonia territoriale”, su cui si è costruita la storiografia analogica, ma per utilizzarle come filo conduttore di una analisi computazionale che vuole colmare quel vuoto determinato non già da un errore metodologico, quando della limitata possibilità d'analisi legata alla ricerca non informatizzata.

2. L'organizzazione del database

L'archivio della famiglia Paternò Castello, principi di Biscari (oggi, Acate – RG), conservato presso l'Archivio di Stato di Catania, rappresenta un prezioso complesso di dati storici operazionabili, la cui eterogeneità può rispondere con particolare efficacia ad un approccio metodologico in ambiente digitale, anche nella prospettiva di costruzione dei Big Data della Storia.

Questo patrimonio, infatti, proprio perché espressione di una dinamicità sociale, culturale e politica non già del suo soggetto produttore, quanto di una comunità che, insieme a lui, mostra le riflessioni dell'uomo verso il suo territorio, all'interno di un processo di digitalizzazione che non sia la semplice restituzione fotografica delle carte, può diventare banco di prova e laboratorio epistemologico volto ad arricchire quella storiografia che ha messo Catania al centro del suo interesse – così come già il progetto “Venice Time Machine” ([1];[31]) ha avuto modo di dimostrare per la Serenissima.

Tra i 2.000 mila pezzi custoditi, che registrano politiche, economie, diatribe e controversie, scelte governative e posizioni di potere, la sola sezione “Corrispondenza” consta di circa 42.493 carte, nella quale, alla corda 1642, ne troviamo 591. Proprio queste, vanno a costituire 366 lettere e 1

manoscritto di Emile Rousseau.

Anche se per un approccio quantitativo sarebbe stato opportuno scegliere delle fonti quali, ad esempio, i *registri delle entrate e delle uscite* (tra l'altro, numerosi nell'archivio), la selezione della corda è stata determinata, da un lato, dalla sua eterogeneità, e, da un altro lato, da una presa di posizione di campo che mira, in ambiente informatico, alla riconfigurazione del quadro dei rapporti della famiglia Paternò Castello; un approccio la cui efficacia, in termini di descrizione e analisi, non può non tener conto delle prospettive aperte da Ruth e Sebastian Ahnert, e Kim Albrecht, confluite nel "Tudor Network" (<https://tudornetworks.net>).

Il database è stato realizzato su soluzione software Claris Filemaker Pro 19 – scelta dettata dal fatto che tale tool non prevede, data la fase iniziale del lavoro, la condivisione su server –, e l'intera GUI è stata strutturata per consentire la consultazione del documento fotografato e la possibilità di effettuare ricerche in esso.

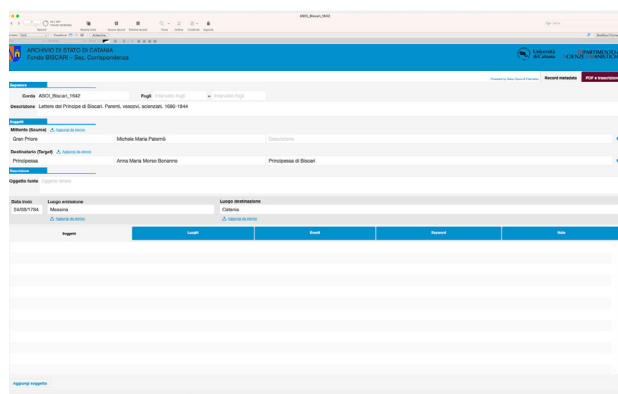


Figura 1. Interfaccia del record sul database realizzato con Filemaker.



Figura 2. Interfaccia del PDF.

Ogni singolo record raccoglie le immagini che vanno a costituire le singole lettere, corredate da tutti quei metadati necessari ad una successiva analisi: mittente, destinatario, data e luogo di emissione, luogo di destinazione, soggetti e luoghi nominati, regesto e keywords.

Ogni singola informazione è indicizzata, ed i 367 record sono stati costruiti sulla base di una sottostruttura relazionale del database, dove vengono inseriti i singoli soggetti individuati, ed i vari luoghi, anch'essi in sottosistema – quest'ultimi corredate di coordinate GPS, per consentire di collocare i vari documenti e gli attori in una mappa digitale.

I due sottosistemi consentono di comprendere come questi attori si relazionavano gli uni con gli altri e dove collocavano i loro interessi. Il record, infatti, viene compilato mediante fasi di importazioni dei dati, allo scopo di realizzare un singolo “contenitore” di informazioni correlate. Questo meccanismo consente, successivamente, di esportare dati in vari formati (ad esempio, .csv), con la possibilità, soprattutto, di tener conto delle informazioni che vogliono ricavarsi (la sola relazione tra mittente-destinatario, quella spaziale o temporale, o tutte insieme). Ad esempio, “The Vistorian” (<https://vistorian.net>) è stato utilizzato su un file .csv contenente solamente quei dati utili per visualizzare il network “sender-consignee”; mentre, l’elaborazione di un diverso file, su “Palladio” (<https://hdlab.stanford.edu/palladio/>), ha restituito una rete che presenta le stesse caratteristiche di quelle ricavate da The Vistorian, consentendo, però, di geolocalizzarla, mostrando come si estendeva nello spazio quell’intreccio di relazioni.

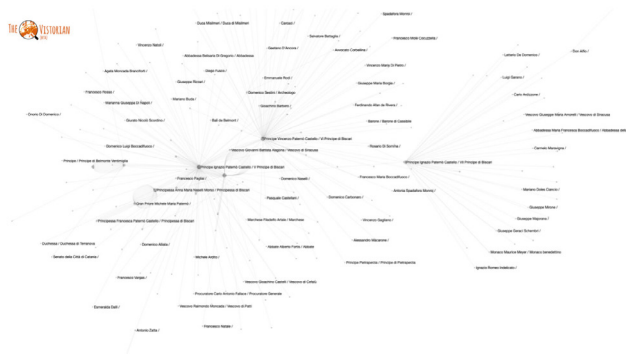


Figura 3. Rete dei soggetti su “The Vistorian”

I centri della gestione di questo potere nobiliare sono legati alle “grandi capitali”, Catania, Palermo e Napoli, ma il network abbraccia la maggior parte della Sicilia orientale (fig. 4), per spingersi oltralpe (Zurigo, Vienna) e oltremare (Malta).

Seppur distanti nel tempo, le lettere dimostrano quale considerazione vi era, nell’orizzonte internazionale, dell’operato dei principi e delle principesse di Biscari, e quale profonda capacità politica riusciva ad esprimere questa casata nobiliare. Da Vienna, il 4 aprile 1786, il principe di Belmonte Ventimiglia scrive ad Ignazio (il 5° principe) per chiedergli di curare e aiutare il Conte Fise (Fisc?) affinché il suo *grand tour* potesse essere vissuto con tutto «il riguardo» che Ignazio era solito dare a chi si avvicendava a conoscere la Sicilia e le sue ricchezze. Ancora, l’1 dicembre 1830, da Zurigo, «la métropole du Zvinglianisme»², Padre Maurice Meyer, dell’ordine de S.

2 ASCt, Archivio Biscari, *Fondo corrispondenza*, v. 1642, ff. 224r-225v.

Benoit, chiede aiuto a Ignazio (7° principe) per salvare cinque bambini dalla “minaccia” del credo protestante: la comunità calvinista, non riconoscendo il diritto di dimora ai cattolici, obbligava questi bambini al trasferimento presso una località della Svizzera francese. L’unico rimedio sarebbe stato l’acquisto di tale diritto presso una comunità cattolica, ma al costo di 1.200 franchi. Meyer chiede, così, ad Ignazio una donazione di 800 franchi per consentire a questi bambini di non essere allontanati e non essere obbligati alla conversione verso il protestantesimo. Mentre è alla principessa Francesca Paternò Castello Arezzi (moglie di Vincenzo, 6° principe) che – da Malta, il 26 marzo 1793 – Bali de Belmont invia una missiva con la richiesta di un’intercessione presso il Senato di Catania, per trovare una soluzione alla crisi frumentaria che in quel momento storico attanagliava il mercato mediterraneo, in crisi a causa delle pressioni ideologiche e dei blocchi scaturiti dalla Rivoluzione Francese.

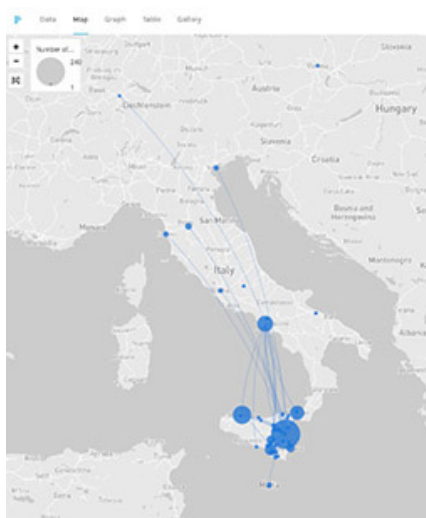


Figura 4. Network dei luoghi e la sua distribuzione su mappa georeferenziata

Quella dei Paternò Castello è un’ascesa che si dipana nei secoli – è nel 1623 che Agatino riceve, da Filippo 4° di Spagna, il titolo di «principe» ([43]) –, che raggiunge la sua massima espressione politico-culturale con la quinta, sesta e settima titolatura, ossia con Ignazio e, successivamente, col figlio Vincenzo e, ancora, con il nipote, che eredita il nome del “grande” (fig. 5). Il network si espande e si diversifica, mostrando come gli aspetti politici diventino sempre più profondi, ma che, proprio con Ignazio, si arricchiscono di quel lato culturale che non è semplice ostentazione o mecenatismo d’opportunità, bensì espressione di una forte personalità culturale intenta a ricostituire il patrimonio della città di Catania. Per questo, il principe intesse una fitta rete di relazioni con la comunità scientifica locale ed europea, da un lato per portare la sua casata dentro il quadro culturale italiano ed europeo, e, da un altro lato, per avvalersi delle più autorevoli voci in grado di supportarlo nella raccolta di tutte quelle ricchezze che, progressivamente, andavano a formare uno dei più importanti patrimoni d’antiquaria siciliani.



La rete al 1774

La rete al 1624



La rete al 1840

Figura 5. Espansione del network 1624-1840

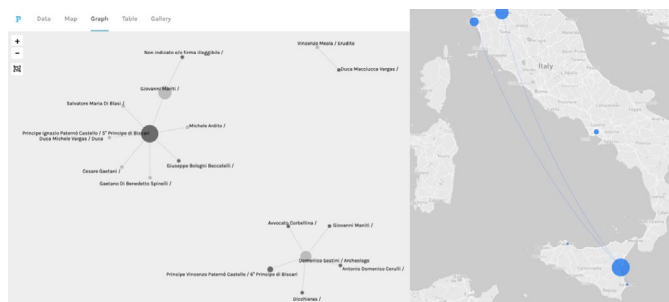


Figura 6. Il network degli scienziati

Il network scientifico, invece, si estendeva, per lo più, sul territorio della penisola italiana, con *nodes* principali in Toscana (fig. 6). Ogni scienziato, studioso e filosofo era chiamato ad approfondire la storia della Sicilia, focalizzando l'attenzione su svariate questioni; si va dall'analisi di incisioni su marmi che erano stati ritrovati, alla corretta collocazione di statue e colonne su spazi urbani, dalle *plumbate* ai frumenti selvatici, finanche la disputa sulla fondazione e distruzione di città siciliane – da specificare che il filo conduttore del set di dati resta la cultura classica e lo studio dei reperti che Ignazio trovava nei suoi lavori di scavo.

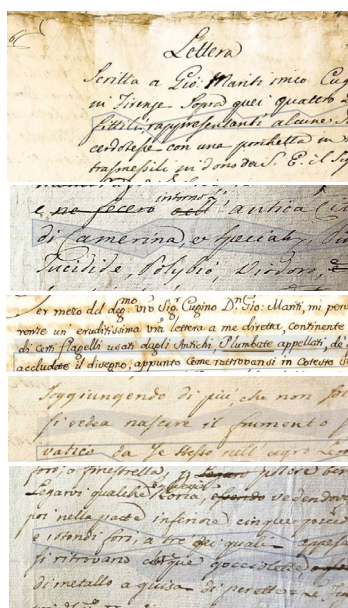


Figura 7. Segmentazione del testo delle fotocopie nella piattaforma Transkribus.

Le lettere – tra le evidenze che emergono – sono per lo più delle copie che i destinatari spedivano al principe, il quale era interessato alle corrispondenze tra gli studiosi stessi, in merito a specifiche tematiche. È così, ad esempio, il caso della lettera di Vincenzo Meola inviata al duca Macciucca

Vargas, in cui si espone il parere scientifico in merito ad una iscrizione greca su di un marmo³.

Tale aspetto è ancor più evidente nello scambio epistolare con Domenico Sestini, l'archeologo che Ignazio volle alla cura del suo museo. La relazione tra i due viene alimentata, infatti, da interposta persona, Giovanni Mariti – cugino e maestro di Domenico –, a cui, così come emerge, spetterà il compito di gestire tale corrispondenza, informare il principe sulle attività dello studioso, sui suoi spostamenti e, soprattutto, delle sue relazioni con il Medioriente⁴. È Mariti, quindi, l'uomo al centro dell'amicizia tra Sestini e Ignazio, ed è ancora lui che si farà tramite del legame scientifico tra Ignazio e l'Accademia dei Georgofili, che riconosceranno a quest'ultimo il titolo di "socio corrispondente"⁵.

Il network sociale del principe si compone, ancora, di altri uomini e studiosi, a testimonianza degli interessi e della variegata personalità culturale di Ignazio: Cesare Gaetani, Salvatore Maria di Blasi, Gaetano di Benedetto Spinelli, Michele Ardito, Francesco Vargas, il duca Michele Vargas, Giuseppe Bogni Beccatelli.

3. Il carteggio Morso-Paternò

Uno degli aspetti fondamentali della Digital History è la sua possibilità di narrare "visualizzando" (*visualization*), ossia attraverso un prodotto di design ([6]), il quale consente di guidare lo studioso verso una visione complessiva dell'oggetto di studio. E sicuramente, tra gli approcci più utili e frequentemente utilizzati, quello della Historical Network Analysis (HNA) ([3];[4];[8];[24];[28];[37]) risponde in maniera decisiva alla necessità dello storico di dare un senso alle informazioni, creare nuova conoscenza ([27]), corroborare teorie e iniziare quel percorso interdisciplinare tra metodi computazionali, indagine umanistica e compilazione di un prodotto editoriale finale.

La HNA si fonda sull'assunto che l'uomo costruisce relazioni, sia con altri simili, che con lo spazio circostante (anche in assenza di altri soggetti). E questo principio, in una prospettiva di realizzazione di un design dell'informazione, si traduce nell'opportunità, per i digital historians, di ricostituire visivamente, in uno schema aperto che fornisce informazioni diverse in base all'oggetto della ricerca, «how humans interpret the world around them» ([4]).

3 *Ivi*, ff. 220r-223v.

4 *Ivi*, ff. 203r-204v.

5 *Ivi*, f. 198r.

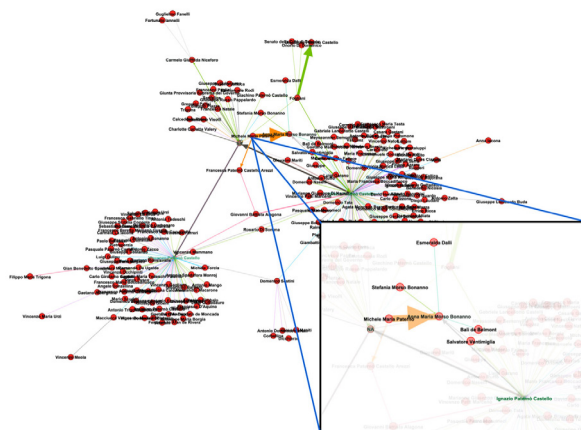


Figura 8. Il carteggio Michele Maria Paternò – Anna Maria Morso e Bonanno su “Gephi”

Così, tra *node e edge*, nella fattispecie, il network restituito da “Gephi” (<https://gephi.org>) (fig. 8) ha consentito di visualizzare una particolarità: all’interno della corda scelta, ben 28 lettere fanno riferimento, quale mittente, al priore di Messina, Michele Maria Paternò ([11]). Il numero significativo, rispetto al trend di tutto il set – per lo più costituito da singole lettere o da poche unità –, ha permesso un’ulteriore analisi mediante il web-based tool “Keyphrase Digger” (fig. 9): il Priore è un uomo nel gioco del sistema politico siciliano, con forti legami con Palermo e Napoli; ma, soprattutto, è un uomo della principessa Anna Morso, destinataria costante di tali carte, la quale si rivela figura centrale nelle relazioni politiche e amministrative che il casato costruisce sul territorio del regno.



Figura 9. Il carteggio Michele Maria Paternò – Anna Maria Morso e Bonanno analizzato con “Keyphrase Digger”

Quello di Anna è un profilo decisamente diverso rispetto a quello del marito Ignazio, la cui attenzione verso l’antiquaria lo rende degno di investiture e riconoscimenti ([42]), mentre Anna ha la necessità di comprendere quello che accade a Palermo e presso la corte di Napoli, soprattutto alla luce di quelle alleanze internazionali che mettevano la Francia al centro del dibattito e la eleggevano protagonista della costruzione di un’asse politico da realizzarsi con la Spagna o con l’Austria; un meccanismo che si sarebbe tradotto nella possibilità di portare le Sicilie sotto

l'influenza asburgica, che, se da un lato era alla ricerca di sbocchi commerciali sul Mediterraneo, dall'altro lato si mostrava come una monarchia sicuramente più razionale ed illuminata rispetto a quella spagnola, che poco aveva giovato, in termini economici e commerciali, al sud dell'Europa – indicativo, da questo punto di vista, il fatto che Carlo III fu sempre ostile all'idea di impegnarsi per sottrarre Napoli e Palermo dalla sudditanza economica della grandi potenze europee ([52]).

Nella seconda metà del Settecento, il quadro internazionale incise, quindi, a fondo sui destini delle Sicilie, spingendole verso il baratro, dove il carattere maturo di Bernardo Tanucci cercò di arginare quello poco accorto, né colto ed erudito, di Ferdinando IV, ma con l'amaro epilogo della sua destituzione e sostituzione con Giuseppe Beccadelli, duca della Sambuca, le cui azioni sul filo del rasoio – un filo austriaco che guardò con interessi anche alla Spagna – spinsero i napoletani ad architettarne la fine. Così, infatti, scrive Michele Maria alla principessa Anna:

«Da Napoli mi si dice, che una mattina si presentò al S. Mar.se della Sambuca un giovane del riposto, il q.le gli disse di non prender cioccolatte, perché aveva veduto mettere un non so che dentro dal ripostiere. Di fatti, il Mar.se non lo prese, ma fattolo bere ad un cagnolino, questi sene morì, ed il Ripostiere si dice carcerato. Se ne parla da per tutto, ma resta ignota la verità di q.o fatto»⁶.

La Sicilia di Anna doveva prepararsi a nuove vedute – probabilmente da evitare –, ma che mettevano i siciliani nella direzione di un cambiamento radicale che avrebbe, per arricchire il «Regio Erario», destabilizzato la dimensione politica e amministrativa di quella terra:

«In Palermo – scrive Michele Maria – vi sono lagnanze universali; per avere il Sovrano abbassato le Soggiogazioni dal cinque al 4 p. 100, e l'un p. 100 depositarsi in Tavola per conto del Regio Erario; che S.E. avendo avanzati 22 progetti per mutare la forma del Governo della Sicilia, la M.S. abbia rimesso lo esame di questi alla Giunta delle Finanze, dove assiste con impegno il consultore Simonetti ivi destinato dal Viceré. Fra questi si vuole l'abolizione del Parlame.to, e della Deputazione del Regno, tolto il Mero misto a' baroni, ed il luogo del Viceré, tre Presidi indipendenti l'uno dall'altro i q.li amministrino giustizia colla loro piccola corte; e che il passaggio delle cause feudali seguisse in Napoli; abolito il Trib.le del Concistoro, e soppressi tutti i Senati da sostituirsi in loro vece l'Eletto del Popolo, e finalmente che sopra ogni Salma di terreno si pagassero tari nuovi, progetti son questi ed altri co' q.ti ha fatto presente alla corte il S. Viceré il vantaggio di più milioni che ne va a ritraere il Regio Erario»⁷.

Il Regno cercava una sua “rivoluzione” – certamente nulla si immaginava di quello che sarebbe accaduto in Francia e in Europa dal 14 luglio 1789 –, e dentro il quadro delle attrici che ebbero un ruolo determinante nella scrittura degli equilibri volti ad evitarla, Anna Maria Morso e Bonanno ebbe il compito di costruire e sostenere le reti politiche che la sua famiglia tessera con gli organi di governo locale e quello al di là del faro; a lei è, così, profondamente legato il priore messinese Michele Maria Paternò, che costantemente “la magnifica” e a cui rivolge il pensiero nei momenti più critici della sua posizione politica, dentro e fuori il suo ruolo. Anna, dal suo canto, non può far a meno di Paternò, attivando tutte le sue forze per tutelarla e seguirlo, finanche abbracciare il destino sociale del suo pupillo – il liberato schiavo Michelino – che diventerà il *casus* degli attacchi al priore ([5]) da parte della sua famiglia d'origine, quando egli perderà il sostegno della principessa Anna, chiamata a pagare, il 4 gennaio 1792, il suo debito con la vita.

6 *Ivi*, f. 383v.

7 *Ivi*, f. 383r.

4. Transkribus e filemaker

Come già evidenziato (vedi *supra*), la GUI del database consente la piena consultazione dei facsimili, ma, a differenza di altri sistemi oggi presenti sul Web, i singoli PDF delle lettere sono stati creati con l'ausilio dell'Intelligenza Artificiale Transkribus, la quale consente di realizzare file che permettono analisi testuali, ma, soprattutto – ed è questo il *focus* più importante di questa tecnologia! – ottenere la trascrizione automatizzata di testi manoscritti.

Quello dell'automatic transcription ha da sempre rappresentato un limite per un reale sviluppo delle metodologie informatiche nella ricerca storica. Già Dimond, nel 1957, metteva in evidenza la necessità della «character-recognition» ([13]) per i documenti archivistici, con la finalità di creare testi «into forms that computers can process».

I progressi in ambito statistico e i recenti sviluppi delle reti neurali, per potenziare i sistemi già in dotazione al mondo finanziario e postale, per il riconoscimento dei caratteri degli indirizzi postali scritti sulle buste ([46]), o i dati calligrafici degli assegni bancari ([12]), ha aperto il varco per una ulteriore applicazione delle Intelligenze Artificiali ai documenti storici.

Già dal 2018, progettato da Adam Matthew Digital, “Quartex” ambisce ad essere vertice dei servizi di digitalizzazione (<https://www.quartexcollections.com>) di documenti manoscritti, anche se il servizio è riservato ad organizzazioni di ricerca, impedendo a studiosi indipendenti di potenziare l'IA su cui si basa.

Parallelamente a tale ambito di ricerca e sviluppo, si è posta all'attenzione della comunità degli storici la categoria di “Big Data”, che proprio in questo settore di studi trova la sua valida concettualizzazione in riferimento alle fonti primarie native digitali, o a fonti dattiloscritte o a stampa (analogica), su cui inizia la riflessione sulla possibilità di applicare sistemi di lettura automatica dei caratteri (OCR).

La costruzione di database e archivi digitali, infatti, se da un lato rappresenta una frontiera tecnologica valida per l'avanzamento delle ricerche storiche, da un altro lato serve solamente a organizzare immagini e facsimili, ma non a consentire analisi testuali direttamente sulle immagini acquisite. Perché si possano processare, anche i manoscritti devono essere digitalizzati, ma con una specifica ulteriore: il testo in esso contenuto deve essere codificato e tradotto in un formato *machine readable* (ad esempio: file “.txt”).

I Big Data della Storia, quindi, devono costruirsi; ed il passo più importante è quello di una digitalizzazione che sia reale costruzione di un patrimonio di informazioni processabili, un *corpus* digitale che possa soddisfare le necessità computazionali, semantiche, linguistiche e geo-spaziali ([18];[23];[34];[49]) della comunità scientifica. Un paradigma, questo, che si traduce, tra il 2015 e il 2019, nel Transkribus project, «a functioning online research infrastructure where new technologies can feed innovation in archival research» ([41]).

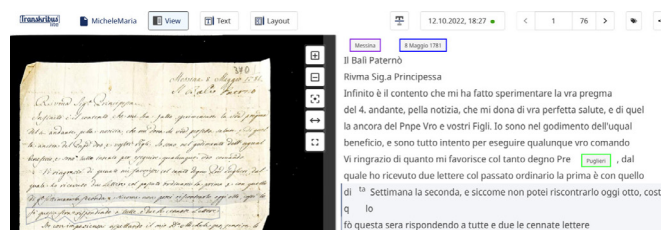


Figura 10. Interfaccia di “Transkribus Lite”.

Basato su Java, Transkribus consente di creare flussi di lavoro imperniati su “reti neurali profonde” ([1];[8]-[10];[21];[25];[28];[31]-[33];[42];[48];[50]) le quali, sulla base di un processo di *pairing* con dei modelli (pubblici)⁸ già presenti nell’Intelligenza Artificiale, restituiscono la trascrizione di quelle parole riconosciute grazie all’addestramento.

Oltre agli strumenti messi a disposizione da Transkribus, all’utente è data anche la possibilità di costruire un proprio modello che formerà in completa autonomia, o, ancora, di effettuare un *boost* su un modello esistente, attraverso un addestramento che richiede l’annotazione manuale di un numero di pagine adeguato all’ammontare di immagini di testo: un «ground truth» di 15.000 parole trascritte (o circa 75 pagine) è generalmente sufficiente per addestrare un modello a riconoscere il testo scritto dalla stessa mano.

Nel nostro caso, la trascrizione automatica del corpus digitale è avvenuta senza alcuna attività di training e modellazione, ma è stato possibile ottenere una trascrizione con una bassa percentuale d’errore, utilizzando il modello pubblico “Italian Administrative Hands 1550-1700”⁹, il cui CER è già del 9,15%. Questa percentuale rappresenta un ottimo risultato per un modello pubblico da applicare per la trascrizione automatica dei manoscritti conservati negli archivi italiani. Questi ultimi, infatti, custodiscono il più vario patrimonio storico d’Europa, non solo per i diversi stili scritturali ma anche per la presenza di diverse lingue e istituzioni (ad esempio il latino in Vaticano, lo spagnolo nel Regno di Sicilia e il Ducato di Milano, l’“italiano” del Granducato di Toscana, ecc.).

Grazie all’IA di Transkribus, i vari PDF inseriti nel database consentono una consultazione con funzionalità pari a quella dei file digitali nativi, ma a questa si aggiungano: (1) la possibilità di esportare il testo “riconosciuto” in diversi formati, che consentono allo studioso di utilizzare ulteriori strumenti per la loro analisi; (2) l’integrazione con la piattaforma Filemaker ed i suoi meccanismi di ricerca nei “campi contenitore”, la qual cosa risponde all’idea gardiniana ([19]-[22]) che la costruzione di database (e, successivamente, di specifiche ontologie) non può prescindere dalla fase di “codifica e formalizzazione” ([41];[42]).

8 L’IA Transkribus è stata sviluppata in seno al progetto “Read” UE di Horizon 2020, da un consorzio di gruppi di ricerca provenienti da tutta Europa, guidato dall’Università di Innsbruck (Austria). Link: < <https://readcoop.eu> >, ultima consultazione 20/10/2022.

9 Il modello è stato addestrato da Jake Dyble, Antonio Iodice, Sara Mansutti e Rachel Midura. Link: < <https://readcoop.eu/model/italian-administrative-hands-1550-1700/> >, ultima consultazione 20/10/2022.

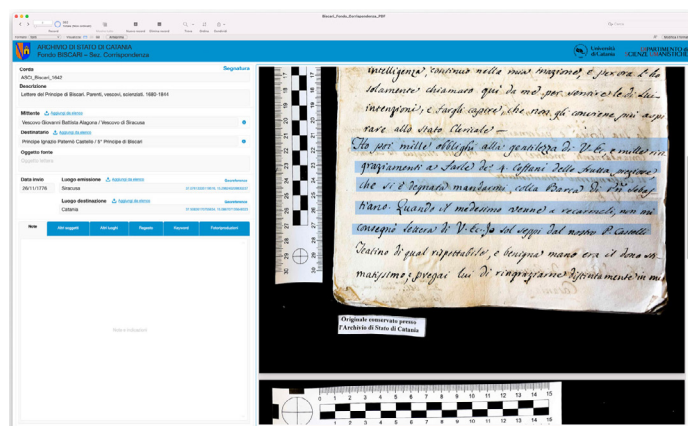


Figura 11. Integrazione PDF trascritti automaticamente con piattaforma Transkribus Lite e Claris Filemaker

Dal canto suo, con la release della versione 19.5 di Filemaker, è stata implementata una nuova funzione JSON, che consente di differenziare i numeri e il testo, e attraverso l'uso di librerie di sistema, effettuare anche il riconoscimento di segni scrittori da una immagine inserita in un campo contenitore, sia in OCR che HTR – una possibilità che, oltre a riflettere la necessità di far fronte alla digitalizzazione di fonti fotografate, rende i database pronti alla realizzazione di complessi *machine readable*.

Il database che è stato progettato, quindi, cerca di rispondere alle necessità metodologiche della Digital History, focalizzando l'attenzione sull'aspetto più delicato ed importante dell'analisi computazionale, ossia la possibilità di lavorare su facsimili che possano rappresentare dei "manoscritti digitali" – i quali, senza una corretta codifica, restano dei semplici file immagine vuoti e privi di significati –, e la costruzione di dataset che possano integrare, grazie a *tools* come "Palladio", "Gephi" o "Keyphrase Digger", l'analisi compiuta dallo studioso, consentendogli di individuare delle informazioni e relazioni a cui difficilmente si potrebbe pervenire – se non dopo un lungo studio individuale – con un metodo prettamente analogico.

5. Considerazioni

Il termine «digitalizzazione» è entrato con pervasività nel vocabolario della nostra società. Tale concetto, però, viene interpretato, per lo più, in maniera erronea dagli *analogue outsiders*; da un lato perché si mettono sullo stesso piano tre concetti in realtà differenti (*digitization*¹⁰, *digitalization*¹¹, *digital transformation*¹²), e da un altro lato perché gli umanisti "analogici" che cercano di approcciarsi all'ambiente digitale, lo scrivono e lo significano quale sinonimo di «facsimile» da utilizzare su un monitor, senza considerare che il processo di digitalizzazione è ben altra cosa che il semplice atto di fotografare/scansionare un documento d'archivio; senza tener in considerazione che se è vero che un documento è un complesso di informazioni, è ancor più

10 ... è la conversione in digitale di una informazione analogica.

11 ... la creazione nativa di informazioni in formato digitale.

12 ... il processo che mira all'adozione esclusiva di tecnologia nei processi di produzione.

vero che è proprio la digitalizzazione a consentirci di creare dei testi che, partendo dalla versione analogica, ci fa pervenire a quella in cui i dati possono realmente soddisfare il quarto paradigma di Gordon Bell: portaci verso una ulteriore conoscenza.

Nel nostro caso, l'automatic transcription – che ha consentito di ottenere rapidamente la versione digitale dei testi – e l'uso di strumenti per l'analisi computazionale, hanno permesso di sviluppare un percorso metodologico che dalla network analysis, è riuscito a portare l'intuizione ad un livello più profondo.

Reti di persone, luoghi ed eventi diventano, difatti, il background necessario per una ontologia unica; ma step necessario resta, inconfutabilmente, la possibilità di lavorare su testi digitali, attraverso l'automazione della trascrizione delle fonti manoscritte. Ciò perché al calcolatore sfugge l'informazione "presente" nell'immagine, che resta, fondamentale, un insieme di pixel che, sommati, raggiungono una certa dimensione, in termini di byte, di definizione.

Quel facsimile rappresenta, per noi, qualcosa di ben diverso. È ciò che legge lo studioso, infatti, che rende quell'immagine (documento) essenziale alla ricostruzione di un evento del passato, una informazione storica, ma per il computer è solo un file silenzioso.

Ecco perché il fondo Biscari è un patrimonio essenziale la cui potenzialità in termini di dati ed informazioni resta ancora tutta da esplorare. Ed il progetto riveste le caratteristiche di un invito a guardare a tale patrimonio archivistico – e a tutti gli altri – sotto i corretti paradigmi della Digital History e delle tecnologie informatiche; a guardare, quindi, ad un maturo investimento nella digitalizzazione che, seppur ancora nel pieno della "fase dell'acquisizione" di tutto quello che l'uomo ha creato fino ad oggi, deve razionalizzare il fatto che il mondo digitale richiede una configurazione specifica che è oltre la semplice acquisizione fotografica.

Riferimenti bibliografici

- [1]. Alison Abbott. 2017. The 'time machine' reconstructing ancient Venice's social networks. *Nature* 546 (2017), 341–344.
- [2]. Hamed Aghdam and Elnaz Heravi. 2017. *Guide to Convolutional Neural Networks. A Practical Application to Traffic-Sign Detection and Classification*. DOI:<https://doi.org/10.1007/978-3-319-57550-6>
- [3]. R. Ahnert, S. E. Ahnert, Y. Ryan, and Computational Humanities Research (CHR 2020). 2020. Networking Archives: Quantitative History and the Contingent Archive. In *Proceedings of the Workshop on Computational Humanities Research (CHR 2020)*, 397–419. Retrieved July 8, 2022 from <https://qmro.qmul.ac.uk/xmlui/handle/123456789/68412>
- [4]. Ruth Ahnert, Sebastian E. Ahnert, Catherine Nicole Coleman, and Scott B. Weingart. 2020. *The Network Turn. Changing Perspectives in the Humanities*. Cambridge University Press. Retrieved July 8, 2022 from <https://www.cambridge.org/core/elements/network-turn/CC38F2EA9F51A6D1AFCB7E005218BBE5>
- [5]. Maria Concetta Calabrese. 2016. Il Gran Priore dell'Ordine di Malta e il suo servo. Un episodio di schiavitù mediterranea nel XVIII secolo. *Nuova Riv. Stor.* 100, 3 (2016), 907–976.

- [6]. Stefano Capezzuto. 2017. Il design della conoscenza. Intervista a Jeffrey Schnapp. *il lavoro culturale*. Retrieved November 29, 2018 from <http://www.lavoroculturale.org/intervista-a-jeffrey-schnapp/>
- [7]. Kumar Chellapilla, Sidd Puri, and Patrice Simard. 2006. High Performance Convolutional Neural Networks for Document Processing. Suvisoft. Retrieved March 11, 2022 from <https://hal.inria.fr/inria-00112631>
- [8]. Vincent Chollier. 2018. From court society to power networks. Evolution of social structures in Upper Egypt during the New Kingdom. In *Social Network Analysis in Ancient Egypt*, Prague, Czech Republic. Retrieved July 8, 2022 from <https://halshs.archives-ouvertes.fr/halshs-01947723>
- [9]. Dan C Ciresan, Ueli Meier, Jonathan Masci, Luca M Gambardella, and Jürgen Schmidhuber. 2011. Flexible, High Performance Convolutional Neural Networks for Image Classification. *Proc. Twenty-Second Int. Jt. Conf. Artif. Intell.* (2011), 6.
- [10]. Ronan Collobert and Jason Weston. 2008. A unified architecture for natural language processing: deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning (ICML '08)*, Association for Computing Machinery, New York, NY, USA, 160–167. DOI:<https://doi.org/10.1145/1390156.1390177>
- [11]. Fabrizio D’Avenia. 2000. Le commende gerosolimitane nella Sicilia moderna: un modello di gestione decentrata. (2000), 35–86.
- [12]. G. Dimauro, S. Impedovo, G. Pirlo, and A. Salzo. 1997. Automatic Bankcheck Processing. A New Engineered System. *Int. J. Pattern Recognit. Artif. Intell.* 11, 04 (June 1997), 467–504. DOI:<https://doi.org/10.1142/S0218001497000214>
- [13]. T. L. Dimond. 1957. Devices for reading handwritten characters. In *Papers and discussions presented at the December 9-13, 1957, eastern joint computer conference: Computers with deadlines to meet (IRE-ACM-AIEE '57 (Eastern))*, Association for Computing Machinery, New York, NY, USA, 232–237. DOI:<https://doi.org/10.1145/1457720.1457765>
- [14]. Jack Dougherty and Kristen Nawrotzki. 2016. *Writing history in the digital age*. University of Michigan Press, Michigan.
- [15]. Johann Gustav Droysen. 1994. *Istorica: lezioni di enciclopedia e metodologia della storia, 1857*. Guida, Napoli.
- [16]. Luciana Duranti. 2006. Verso una teoria archivistica di conservazione digitale, i risultati concettuali del progetto InterPares. *Archivi I*, 1 (2006), 75–97.
- [17]. John Alexander Edwards III. 2007. Easily adaptable handwriting recognition in historical manuscripts - ProQuest. University of California, Berkeley. Retrieved February 24, 2022 from <https://www.proquest.com/openview/eff875e0882f5fd341be07b7992145c8/1?pq-origsite=gscholar&cbl=18750>
- [18]. François Furet. 1981. Il quantitativo in Storia. In *Fare storia. Temi e metodi della nuova storiografia*, Jacques Le Goff and Pierre Nora (eds.). Einaudi, Torino.
- [19]. Jean-Claude Gardin. 1959. Four codes for the description of artefacts. An essay in archaeological technique and theory. *Am. Anthropol.* 60, (1959), 335–357.
- [20]. Jean-Claude Gardin. 1960. Les applications de la mécanographie dans la documentation archéologique. *Bull. Bibl. Fr.* 5, (1960), 1–16.

- [21]. Jean-Claude Gardin. 1962. De l'archéologie à l'information automatique. *Bull. Euratom* 4, (1962), 25–29.
- [22]. Jean-Claude Gardin. 1991. *Le calcul et la raison. Essais sur la formalisation du discours savant*. Éditions de l'École des Hautes Études en Sciences Sociales, Paris.
- [23]. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press.
- [24]. Martin Grandjean. 2020. A Conceptual Framework for the Analysis of Multilayer Networks in the Humanities. Retrieved July 8, 2022 from <https://halshs.archives-ouvertes.fr/halshs-02650245>
- [25]. Ian Gregory, Christopher Donaldson, Patricia Murrieta-Flores, and Paul Rayson. 2015. Geoparsing, GIS, and Textual Analysis. Current Developments in Spatial Humanities Research. *Int. J. Humanit. Arts Comput.* 9, 1 (2015), 1–14. DOI:<https://doi.org/10.3366/ijhac.2015.0135>
- [26]. Maria Rosaria Guercio. 2019. *Archivistica informatica. I documenti in ambiente digitale*. Carocci, Roma.
- [27]. Anthony J. G. Hey (Ed.). 2009. *The fourth paradigm: data-intensive scientific discovery*. Microsoft Research, Redmond, Washington. Retrieved October 20, 2022 from <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>
- [28]. Jacob Holland-Lulewicz and Amanda D. Roberts Thompson. 2021. Incomplete Histories and Hidden Lives: The Case for Social Network Analysis in Historical Archaeology. *Int. J. Hist. Archaeol.* (November 2021). DOI:<https://doi.org/10.1007/s10761-021-00638-z>
- [29]. D. H. Hubel and T. N. Wiesel. 1968. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 1 (March 1968), 215–243. DOI:<https://doi.org/10.1113/jphysiol.1968.sp008455>
- [30]. Pat Hudson and Mina Ishizu. 2017. *History by Numbers* (2nd ed.). Bloomsbury Academic. Retrieved January 11, 2022 from <https://www.bloomsbury.com/uk/history-by-numbers-9781474294157/>
- [31]. Frédéric Kaplan. 2015. *The Venice Time Machine*. (2015).
- [32]. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, Curran Associates, Inc. Retrieved December 17, 2021 from <https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>
- [33]. Gerard Labuda. 1957. Próba nowej systematyki i nowej interpretacji źródeł historycznych. *Stud. Źródłoznawcze* (1957). Retrieved February 8, 2022 from <https://rcin.org.pl/dlibra/publication/39120/edition/41495>
- [34]. Emmanuel Le Roy Ladurie. 1973. *Le territoire de l'historien*. Gallimard, Paris.
- [35]. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (November 1998), 2278–2324. DOI:<https://doi.org/10.1109/5.726791>
- [36]. Puttatida Mahapattanakul. 2019. From Human Vision to Computer Vision – Convolutional Neural Network (Part3/4). *Medium*. Retrieved March 11, 2022 from

<https://becominghuman.ai/from-human-vision-to-computer-vision-convolutional-neural-network-part3-4-24b55ffa7045>

[37]. Emilie Martinet. 2017. Analysing the Dynamics among the Social Groups and the Mechanisms of Social Promotion in the Provinces in the late Old Kingdom: SNA Methods and New Research Approaches. Libreria Antiquaria Pontremoli, 260. Retrieved July 8, 2022 from <https://halshs.archives-ouvertes.fr/halshs-02907433>

[38]. Masakazu Matsugu, Katsuhiko Mori, Yusuke Mitari, and Yuji Kaneda. 2003. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Netw.* 16, 5 (June 2003), 555–559. DOI:[https://doi.org/10.1016/S0893-6080\(03\)00115-1](https://doi.org/10.1016/S0893-6080(03)00115-1)

[39]. Albert Meroño-Peñuela, Ashkan Ashkpour, Marieke Erp, Kees Mandemakers, Leen Breure, Andrea Scharnhorst, Stefan Schlobach, and Frank Harmelen. 2014. Semantic technologies for historical research. A survey. *Semantic Web* 6, (October 2014), 539–564. DOI:<https://doi.org/10.3233/SW-140158>

[40]. Mirella Mombelli Castracane. 1993. L'Archivistica tra storia e informatica. *Arch. Stor. Ital.* 555 (1993), 1.

[41]. Guenter Muehlberger, Louise Seaward, Melissa Terras, Sofia Ares Oliveira, Vicente Bosch, Maximilian Bryan, Sebastian Colutto, Hervé Déjean, Markus Diem, Stefan Fiel, Basilis Gatos, Albert Greinöcker, Tobias Grüning, Guenter Hackl, Vili Haukkovara, Gerhard Heyer, Lauri Hirvonen, Tobias Hodel, Matti Jokinen, Philip Kahle, Mario Kallio, Frederic Kaplan, Florian Kleber, Roger Labahn, Eva Maria Lang, Sören Laube, Gundram Leifert, Georgios Louloudis, Rory McNicholl, Jean-Luc Meunier, Johannes Michael, Elena Mühlbauer, Nathanael Philipp, Ioannis Pratikakis, Joan Puigcerver Pérez, Hannelore Putz, George Retsinas, Verónica Romero, Robert Sablatnig, Joan Andreu Sánchez, Philip Schofield, Giorgos Sfikas, Christian Sieber, Nikolaos Stamatopoulos, Tobias Strauß, Tamara Terbul, Alejandro Héctor Toselli, Berthold Ulreich, Mauricio Villegas, Enrique Vidal, Johanna Walcher, Max Weidemann, Herbert Wurster, and Konstantinos Zagoris. 2019. Transforming scholarship in the archives through handwritten text recognition. Transkribus as a case study. *J. Doc.* 75, 5 (January 2019), 954–976. DOI:<https://doi.org/10.1108/JD-07-2018-0114>

[42]. Francesco Muscolino. 2015. Il principe di Biscari e il principe di Torremuzza, «i due Dioscuri della passione antiquaria settecentesca». *LANX* 21, (2015), 1–40.

[43]. Carmelo Nicotra. 1977. *Il Carmelo catanese nella storia e nell'arte*. Tipografia Samperi, Messina. Retrieved May 31, 2022 from <https://opac.bncf.firenze.sbn.it/bncf-prod/resource?uri=SBL0585780&v=l&dncnr=1>

[44]. Tito Orlandi. 1990. *Informatica umanistica*. La Nuova Italia Scientifica, Roma.

[45]. Tito Orlandi. 1996. Formalizzazione dei dati, Semiotica e Comunicazione. *Archeol. E Calcolatori* 7 (1996), 1247–1258.

[46]. Umapada Pal, Ramit Kumar Roy, and Fumitaka Kimura. 2012. Multi-lingual City Name Recognition for Indian Postal Automation. In *2012 International Conference on Frontiers in Handwriting Recognition*, IEEE, Bari, Italy, 169–173. DOI:<https://doi.org/10.1109/ICFHR.2012.238>

[47]. Lewis J. Perelman. 1992. *School's Out: Hyperlearning, the New Technology, and the End of Education* (1st edition ed.). William Morrow & Co, New York.

- [48]. Ragav Venkatesan, Baoxin Li. *Convolutional Neural Networks in Visual Computing: A Concise Guide*. Retrieved March 11, 2022 from <https://www.routledge.com/Convolutional-Neural-Networks-in-Visual-Computing-A-Concise-Guide/Venkatesan-Li/p/book/9781138747951>
- [49]. Raffaele Santoro. 2015. *La grande bellezza dei grandi dati*. Archivio di Stato di Venezia, Roma.
- [50]. Salvatore Spina. 2020. *Archivi nell'era delle Digital Humanities, dei Big Data e della Genetica*. Algra, Viagrande (Catania).
- [51]. Salvatore Spina. 2022. *Digital History. Metodologie informatiche per la ricerca storica*. Edizioni Scientifiche Italiane, Napoli.
- [52]. Roberto Tufano. 2015. *La Francia e le Sicilie. Stato e disgregazione sociale nel Mezzogiorno d'Italia da Luigi XIV alla Rivoluzione*. Arte Tipografica Editrice, Napoli.
- [53]. M. V. Valueva, N. N. Nagornov, P. A. Lyakhov, G. V. Valuev, and N. I. Chervyakov. 2020. Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. *Math. Comput. Simul.* 177, (November 2020), 232–243. DOI:<https://doi.org/10.1016/j.matcom.2020.04.031>
- [54]. Martin Weisser. 2015. *Practical Corpus Linguistics. An Introduction to Corpus-Based Language Analysis*. Wiley. Retrieved February 24, 2022 from <https://www.perlego.com/book/993207/practical-corpus-linguistics-an-introduction-to-corpusbased-language-analysis-pdf>
- [55]. Wei Zhang, Kazuyoshi Itoh, Jun Tanida, and Yoshiki Ichioka. 1990. Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Appl. Opt.* 29, 32 (November 1990), 4790–4797. DOI:<https://doi.org/10.1364/AO.29.004790>