

Istituto di Scienze del Patrimonio Culturale

Consiglio Nazionale delle Ricerche

Sede di Napoli

Il riconoscimento automatico di scrittura per documenti storici

RAPPORTO TECNICO

Stilato nell'ambito dell'assegno di ricerca "Definizione delle strategie di mapping delle informazioni; definizione delle strategie per l'estrazione della conoscenza; ricerca e sperimentazione di sistemi di codifica di documenti medievali e moderni in ambiente digitale" del progetto "PON IDEHA – Innovazioni per l'elaborazione dei dati nel settore del Patrimonio Culturale"

VERA ISABELL SCHWARZ-RICCI

1. Introduzione	2
2. Historical Document Processing	2
3. Principi di funzionamento	3
4. Il flusso di lavoro	4
5. Prospettive future	7
6. Datasets e piattaforme	8
7. Riferimenti bibliografici	11
8. Elenco delle figure	12

1. Introduzione

Con l'incremento della digitalizzazione negli ultimi anni di collezioni di beni culturali e con la loro messa in rete si pone in maniera più accentuata il problema dell'accesso a questi materiali, dato che per ragioni di costi il loro corredo di metadati si riduce spesso ai dati essenziali. Nell'ambito dei beni librari e archivistici sono state sviluppate diverse strategie per colmare almeno in parte questa lacuna, ad esempio con la messa a disposizione di indici digitalizzati, con la possibilità per gli utenti di corredare l'oggetto digitale con parole chiave o iniziative di crowd-sourcing per metadattazione o con la trascrizione di collezioni di particolare interesse per la citizen science. Una soluzione al di là di iniziative singole di questo genere potrebbe però offrirla il campo dell'intelligenza artificiale con l'apprendimento automatico (machine learning), che comprende una serie di metodi di particolare interesse per i beni culturali.¹

Così il riconoscimento di pattern (pattern recognition) trova una sua applicazione nella identificazione di oggetti o persone su quadri o fotografie e nell'ambito del patrimonio scritto nel riconoscimento di testo. Per quanto riguarda libri o altro materiale a stampa, invece, il riconoscimento ottico dei caratteri (optical character recognition, OCR) ha una lunga tradizione alle sue spalle e rileva e identifica i caratteri presenti in un'immagine digitale di un testo stampato convertendoli in cifre (codifica di caratteri) in maniera che siano leggibili e elaborabili da un computer. Per quanto riguarda il materiale scritto a mano, si parla invece di riconoscimento di scrittura (handwritten text recognition, HTR). I primi tentativi in questo ambito si basavano similmente a quanto avviene per l'OCR sul riconoscimento dei singoli caratteri: un approccio che promette successo solo in caso di scritture a mano che imitano i caratteri tipografici, con tassativa separazione delle lettere e uniformità dei loro tratti (in caso dell'alfabeto latino chiamati block letters o print writing) o nell'evenienza di scritture concepite, ancora prima della introduzione della stampa a caratteri mobili in Europa, per la facilità di lettura, le scritture librarie. Infatti, i caratteri tipografici ancora oggi maggiormente in uso in Occidente, quelli della cosiddetta Antiqua, sono stati modellati a loro volta su una scrittura libraria del rinascimento italiano. Dato che molte scritture a mano tendono invece a legare i singoli caratteri tra di loro, trasformando così la loro forma (scrittura corsiva), oggi nel riconoscimento di scrittura si adotta largamente il ravvisamento di intere righe, senza una separazione precedente in caratteri o parole (segmentation-free). Entrambi i metodi, OCR e HTR, richiedono comunque la precedente individuazione della struttura della pagina raffigurata.

2. Historical Document Processing

In questa sede ci interessa in particolare il riconoscimento di scrittura per documenti storici (historical document processing) e quindi un riconoscimento offline, mentre la online recognition si occupa del riconoscimento di scrittura dal vivo, effettuato cioè nel momento in cui si scrive, ad esempio su un tablet, su uno smartphone o su una tavoletta grafica. Nel campo della offline recognition si dispone solo dell'immagine del testo senza informazioni (dirette) sulla genesi delle singole lettere, in particolare sulla sequenza dei tratti. Sfide ulteriori costituiscono da un lato la molteplicità delle scritture, il loro grado di corsività e la individualità di singole mani, fattori che

¹ Per uno sguardo d'insieme e una introduzione nel campo del riconoscimento automatico di scrittura si raccomandano i saggi di Muehlberger 2019, Philips 2020 e Lombardi 2020 su cui si basa anche il presente report.

influenzano non solo la forma e grandezza dei caratteri, ma anche parametri come lo spazio tra le parole, l'inclinazione e così via. D'altro canto anche il supporto condiziona le possibilità di successo, se pensiamo ad esempio a un trapelare dell'inchiostro dal lato opposto o alla distanza tra le righe.

Oltre a ciò c'è da considerare che l'Historical Document Processing abbraccia in maniera generica testimonianze scritte a mano di una moltitudine di epoche, di diversa provenienza, lingua e scrittura per cui costituisce un campo davvero vasto. Qui consideriamo questa tecnologia prevalentemente dalla prospettiva di uno storico che si occupi della storia medievale e moderna europea e di testimonianze scritte con l'alfabeto latino, in latino o in altre lingue europee, anche se il funzionamento del riconoscimento di scrittura in sé è indipendente dalla lingua o dall'alfabeto del documento. Fino a poco tempo fa però la ricerca in questo campo si è concentrata per la maggior parte su scritture e manoscritti occidentali.

3. Principi di funzionamento

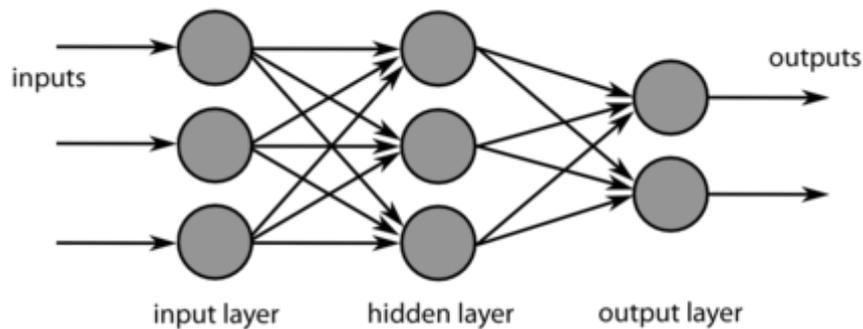


Figura 1: Modello di rete neurale

Dal punto di vista tecnico, per il riconoscimento di scrittura oggi si fa uso di reti neurali artificiali, modelli matematici per il trattamento d'informazione che si ispirano alla struttura del cervello umano con i suoi nodi neurali interconnessi. Le reti consistono di vari strati di nodi che trasmettono segnali allo strato seguente solo se attivati, quindi quando l'input ponderato supera una determinata soglia. Allo stato attuale sono in uso reti neurali ricorrenti in cui i nodi sono interconnessi in un loop o reti convoluzionali in cui i nodi sono connessi in maniera simile alla corteccia visiva. In ciascun strato viene eseguito un riconoscimento di pattern o di caratteristiche visuali e ogni strato lo perfeziona basandosi sull'output dello strato precedente.

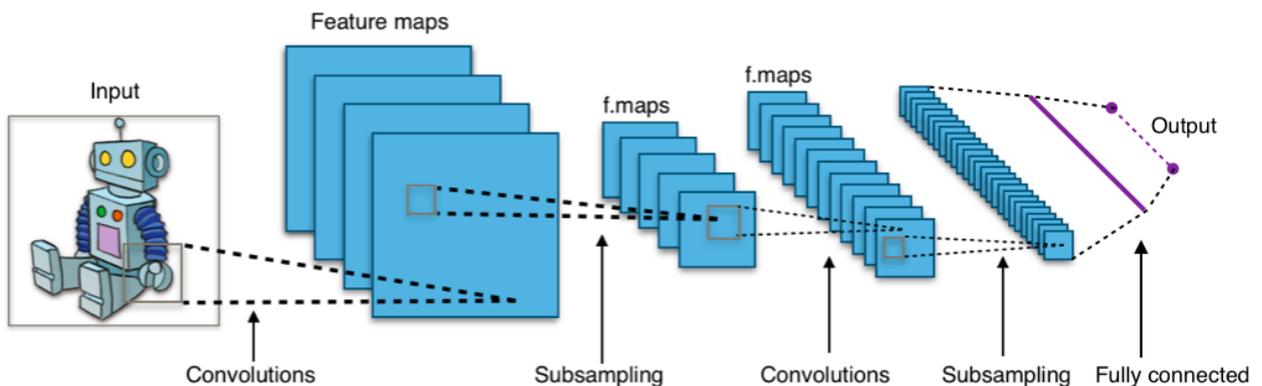


Figura 2: Tipica architettura di una rete neurale convoluzionale

Una rete neurale “deep” consiste di numerosi strati susseguenti ed è in uso nell’apprendimento profondo (deep learning), ma ha bisogno di notevole potenza di calcolo, in particolare un processore grafico prestante.

4. Il flusso di lavoro

Attualmente i sistemi di riconoscimento di scrittura adottano un approccio supervisionato, in altre parole si avvalgono per l’apprendimento di esempi di dataset etichettati o annotati (labeled) ossia di coppie che consistono di un’immagine di riga (input) con la relativa trascrizione (output). Il lavoro con tali sistemi si può suddividere in diverse fasi e qui di seguito descriviamo a titolo esemplificativo un possibile flusso di lavoro.

Dopo aver scelto i manoscritti da trascrivere e averli analizzato per quanto riguarda fattori che influenzano il riconoscimento (stato di conservazione, layout, scritture, lingue e mani presenti) si procede all’acquisizione delle immagini, sia che ciò avvenga tramite la digitalizzazione con mezzi come scanner o macchine fotografiche sia che si passi attraverso il recupero di immagini da biblioteche digitali i cui gestori partecipano all’International Image Interoperability Framework (IIIF)².

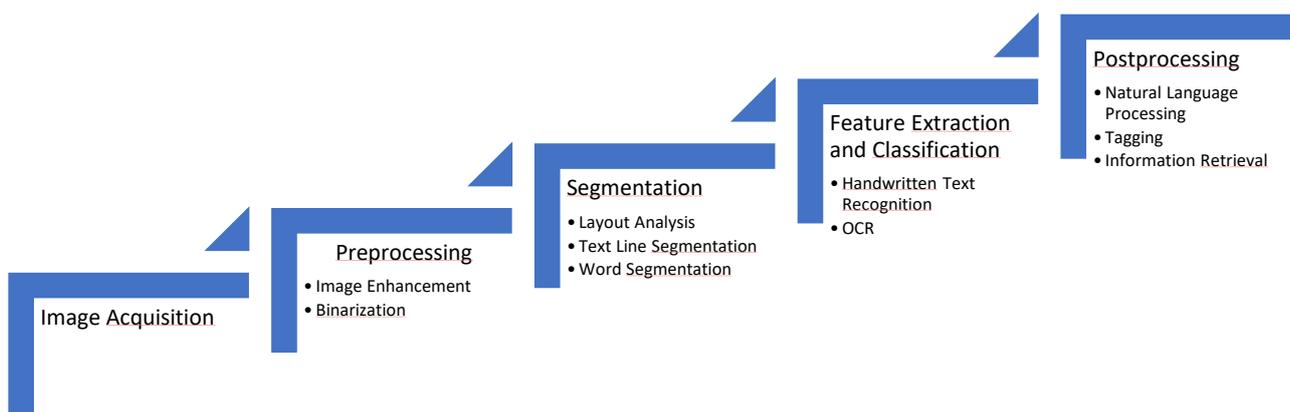


Figura 3: Possibile workflow per il processing di documenti storici

La fase della pre-elaborazione di queste immagini digitali può comprendere il miglioramento delle immagini per facilitare il riconoscimento, ad esempio con la riduzione del rumore o l’eliminazione della distorsione, e la binarizzazione, cioè la trasformazione di un’immagine a colori o a tonalità di grigio in un’immagine a due colori, bianco e nero. La fase seguente, la segmentazione ossia il riconoscimento della struttura della pagina (layout analysis), a volte viene considerata parte della pre-elaborazione. Si tratta della individuazione degli elementi fisici e logici presenti sulla pagina (i primi, ad esempio illustrazioni, il corpo del testo, tabelle; i secondi, titoli, didascalie, note a piè di pagina) con metodi dell’apprendimento automatico (reti neurali convoluzionali, siamesi, U-Nets). Si parte dalla individuazione delle zone in bianco e nero, in altre parole, le regioni di testo, all’interno delle quali si identificano le righe (o anche il paragrafo o le parole). Quest’ultima parte può avvenire

² <https://iiif.io/>. Tutti gli indirizzi web sono stati consultati per l’ultima volta il 6/2/2022.

con una segmentazione in linee di testo o con un riconoscimento delle linee di base (baseline) ossia linea di appoggio delle lettere.

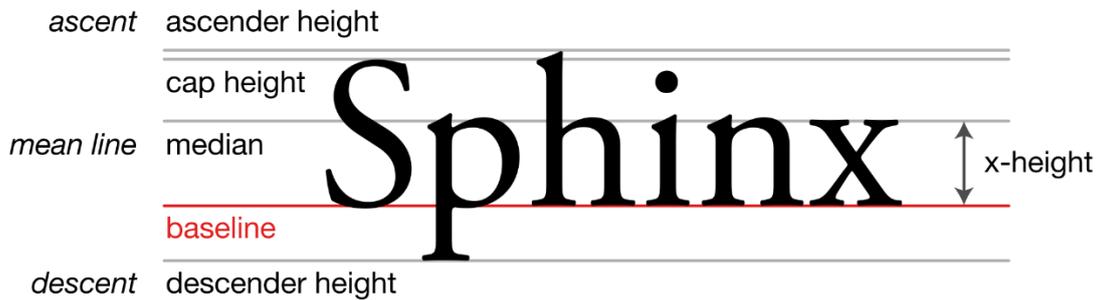


Figura 4: Linea di base

L'importanza di un corretto riconoscimento e posizionamento delle righe e linee non può essere sopravvalutata perché sono essenziali per il passo seguente, il riconoscimento della scrittura, e costituiscono il cosiddetto collo della bottiglia. In caso di documenti storici particolari problemi pongono strutture come righe molte vicine, righe di testo in orientamento non orizzontale, tabelle o iniziali che si estendono per più righe (drop-down initials).



Figura 5: Iniziale dal Cathach di San Columba (sec. VIII)

Il passo seguente è il riconoscimento che richiede un allenamento della rete neurale con un numero rilevante di trascrizioni che vanno inserite in maniera che ogni rigo riconosciuto nell'immagine corrisponda al relativo rigo della trascrizione. L'insieme costituisce la cosiddetta ground-truth o verità di base. Le trascrizioni devono seguire il più fedelmente e omogeneamente possibile i criteri di trascrizioni precedentemente stabiliti per permettere un training di successo. Bisogna tenere presente che la verità di base – anche se una trascrizione è pur sempre un'interpretazione e non può mai riprodurre al 100% la pagina del manoscritto – è comunque la verità fondamentale per la rete neurale la quale trasferisce qualsiasi pregiudizio (bias) presente nella costituzione della ground truth sull'intero set di pagine da riconoscere. Per questo occorre selezionare le pagine per la ground truth con cura e in modo rappresentativo. Per un set di documenti omogenei dovrebbero ammontare a circa 50-100 pagine ossia 15.000 parole per creare un modello (ossia la rete allenata per un determinato set di documenti) con un numero non troppo alto di errori di riconoscimento. Dal numero complessivo della ground-truth si detraggono il 10% per creare il validation o test set e anche qui è d'uopo scegliere delle pagine rappresentative.

Sulla base del training e del validation set si allena un modello la cui qualità si può desumere dalla indicazione delle Word e Character Error Rate (WER e CER). Tra le due la più significativa è l'ultima, dato che una parola può contenere diversi errori al livello di carattere. Nella CER invece confluiscono tutti gli errori, anche quelli che si determinano a livello di spazio e interpunzione, i quali di norma non influenzano la comprensione del testo trascritto. In genere si considera un modello con una CER

sotto i 10% utile, perché al di sopra di questa soglia l'impegno umano per un eventuale correzione manuale è superiore rispetto a una trascrizione ex novo. Strategie per migliorare la performance del modello possono essere ad esempio il cambio di parametri della rete neurale, l'aggiunta di epoche ossia ripetizioni al allenamento, l'aggiunta di ulteriori pagine trascritte alla ground truth o il post-processing.

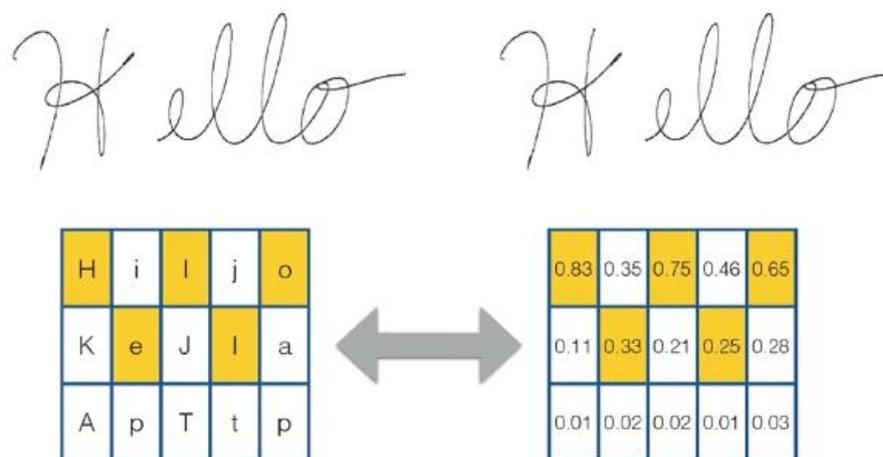


Figura 6: Keyword spotting

In alternativa a queste misure si può con un risultato di bassa qualità anche ricorrere al cosiddetto keyword spotting. Mentre per la trascrizione la rete neurale allenata trascrive per ogni lettera solo quella riconosciuta con la maggiore probabilità, in caso del keyword spotting la rete conserva tutte le lettere possibili con la relativa probabilità e ne costruisce una matrice di probabilità. Per una immagine di una lettera crea quindi una matrice che rappresenta tutte le possibili lettere corrispondenti. Sulla base delle probabilità più alte crea una lista di parole in ordine di probabilità decrescente, recuperando così anche parole con variazioni nella ortografia. Il risultato per l'utente compare non in forma di stringhe di testo, ma piuttosto come evidenziazione della parola nell'immagine, dando in questa maniera la possibilità di giudicare l'esattezza del risultato. Un esempio per questo genere di ricerca lo offre il sito di Himanis Chancery³, che comprende 199 registri della cancelleria reale francese (secc. XIV-XV) e consente una ricerca per parole con caratteri jolly includendo anche la scelta della probabilità desiderata.

³ <http://himanis.huma-num.fr/app/>

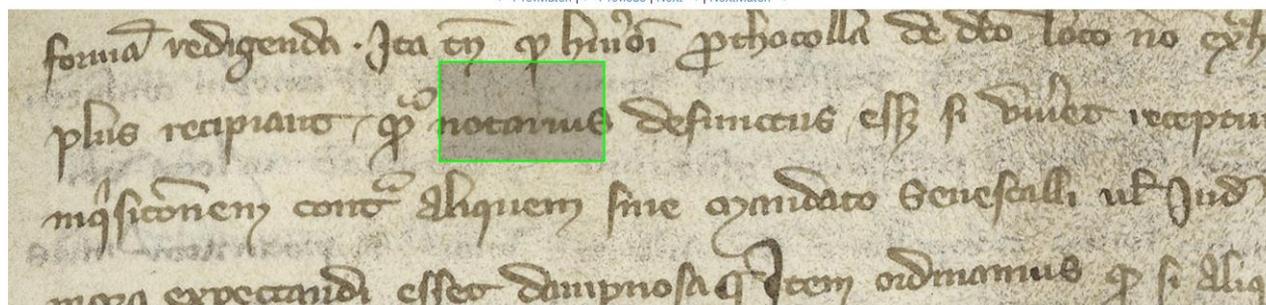


Figura 7: Esempio di keyword spotting in HIMANIS

Da quello che finora abbiamo visto è plausibile che i fattori sopra elencati come lo stato di conservazione, il layout, le scritte e le mani presenti possano influenzare il riconoscimento automatico. Da includere in questo elenco è però anche la lingua. Così come un modello allenato su una determinata scrittura e mano non performerà bene su altra scrittura e mano, un modello allenato su una scrittura in una determinata lingua non funzionerà bene con un'altra lingua perché – anche se nel riconoscimento non è integrato un modello statistico di lingua – la rete impara che determinate sequenze di lettere sono più probabili. Perciò un eventuale riuso di un modello già allenato funzionerà solo con una scrittura simile, la stessa lingua e con l'adozione dei criteri di trascrizione del modello.

Inoltre la lingua gioca un ruolo importante anche per l'ultima fase, il post-processing, nella quale strumenti della elaborazione del linguaggio naturale (natural language processing, NLP) possono essere usati per migliorare la qualità della trascrizione, ad esempio per separare il testo in singole parole (tokenization), per lemmatizzare queste parole e per calcolare la probabilità che determinate sequenze di parole ricorrano nel testo. Al di là del miglioramento del testo esistono anche strumenti per il riconoscimento di entità nominali come nome di persona o di luogo (named entity recognition, NER) che possono aiutare al recupero di informazioni (information retrieval). Nel post-processing può rientrare anche un mark-up, ad esempio per codificare le parti sciolte di abbreviazioni con l'obiettivo di migliorare la performance del modello (Riley 2021). Questo passo può avvenire in maniera automatica o anche con lavoro umano.

5. Prospettive future

Anche se si sostiene a volte che i problemi dal punto di vista della ricerca per il riconoscimento automatico di scrittura siano largamente risolti, questo vale soltanto se il numero delle trascrizioni costituenti la ground-truth è abbastanza alto. Infatti rimangono da districare problemi come il riconoscimento di tabelle, di formule matematiche o di iniziali miniate, ma anche la riduzione della quantità di trascrizioni necessarie per un allenamento o per controllare e migliorare un riconoscimento nel post-processing, in breve, la riduzione della quantità di lavoro umano esperto costoso. La creazione di dati sintetici per l'allenamento è un primo promettente passo verso una soluzione in un campo nel quale i dataset sono notoriamente molto piccoli e eterogenei (Journet 2017). D'altro canto proprio il coinvolgimento umano sembra poter portare a delle soluzioni, in larga misura in progetti che usano non-esperti per la disambiguazione di lettere difficili da riconoscere

(crowdsourcing; Firmani 2018) o in misura più ridotta in progetti che si avvalgono in maniera minima di esperti per un miglioramento delle prestazioni della rete neurale misurando, ad esempio, la percezione umana per rilevare per la rete le parti difficili da trascrivere (Grieggs 2020) o introducendo sistemi assistiti per la trascrizione (Parziale 2021).

Altri campi di ricerca sono il riconoscimento di una pagina per intero (unsegmented), l'allenamento semi- o non-supervisionato, o l'allenamento di reti neurali che svolgono due compiti in contemporanea, ad esempio un simultaneo rilevamento di informazioni linguistiche e di layout. In questa maniera si riduce il rischio del sovradattamento al set di allenamento (overfitting) che può occorrere in caso di un compito unico. Nella comunità scientifica delle Digital Humanities si promuovono inoltre iniziative per creare ontologie per l'analisi del layout di manoscritti⁴ (SegmOnto, Gabay 2021) e sviluppare criteri di trascrizioni che facilitino trascrizioni automatiche⁵, per creare dataset larghi di ground truth omogenea (gold-standard) per il testing di sistemi di riconoscimento e modelli di larghe dimensioni per mani dello stesso stile e di periodi simili (Hodel 2021).

6. Datasets e piattaforme

Anche se mancano ancora set di dati abbastanza larghi e omogenei, si diffonde negli ultimi tempi comunque la buona pratica di pubblicare non solo i risultati della ricerca in articoli, ma anche i dataset su cui si basano le ricerche in repository istituzionali o generici come ad esempio Zenodo⁶ o github⁷. Inoltre si dispone da tempo di dataset open specifici per testare reti neurali su determinati compiti, ad esempio DIVA-HisDB⁸ con immagini da tre codici medievali provenienti da e-codices⁹ per l'analisi del layout o IAM-HisDB¹⁰. Quest'ultimo comprende tre dataset eterogenei, un manoscritto medievale conservato a San Gallo in latino, manoscritti medievali del Parzival in tedesco e autografi di George Washington in inglese, per scopi sia di Layout analysis sia di riconoscimento di scrittura. Il dataset READ-BAD¹¹ contiene campioni da nove collezioni di materiali archivistici medievali e moderni molto eterogenei per il riconoscimento di scrittura.

Anche se per l'implementazione di una rete neurale esistono diverse librerie software open, ad esempio Tensorflow sviluppato da Google¹² o Pytorch sviluppato da Facebook¹³, ci interessano qui piuttosto le piattaforme che offrono una intera gamma di servizi con un'interfaccia web per utente, in particolare eScriptorium, Transkribus e AMAP.

⁴ <https://github.com/SegmOnto>

⁵ <https://cremmalab.hypotheses.org/seminaire-creation-de-modeles-htr>

⁶ <https://zenodo.org/>

⁷ <https://github.com/>

⁸ <https://diuf.unifr.ch/main/hisdoc/diva-hisdb.html>

⁹ <https://www.e-codices.ch/>

¹⁰ <https://fki.tic.heia-fr.ch/databases/iam-historical-document-database>

¹¹ <https://zenodo.org/record/257972>

¹² <https://www.tensorflow.org/>

¹³ <https://pytorch.org/>

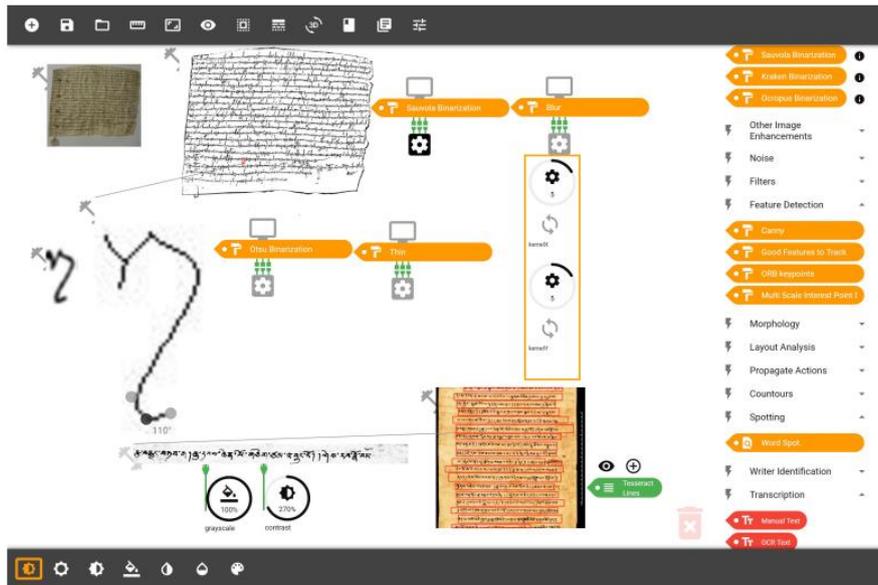


Figura 8: Interfaccia di AMAP

L'Advanced Manuscript Analysis Portal (AMAP, Rajan 2019)¹⁴ mette a disposizione dell'utente una collezione di strumenti compatibili per il pre-processing, la segmentazione e il riconoscimento che possono essere selezionati e combinati usando un linguaggio di programmazione visuale, quindi manipolando elementi grafici con il mouse, in questo caso blocchi che si agganciano in sequenza. Il risultato di ogni manipolazione dell'immagine iniziale appare direttamente nell'interfaccia: Attraverso il simbolo dell'ingranaggio si possono selezionare opzioni per ogni tool. Attualmente il portale sembra però basarsi unicamente sullo sviluppatore e la pagina ufficiale spesso risulta non disponibile.

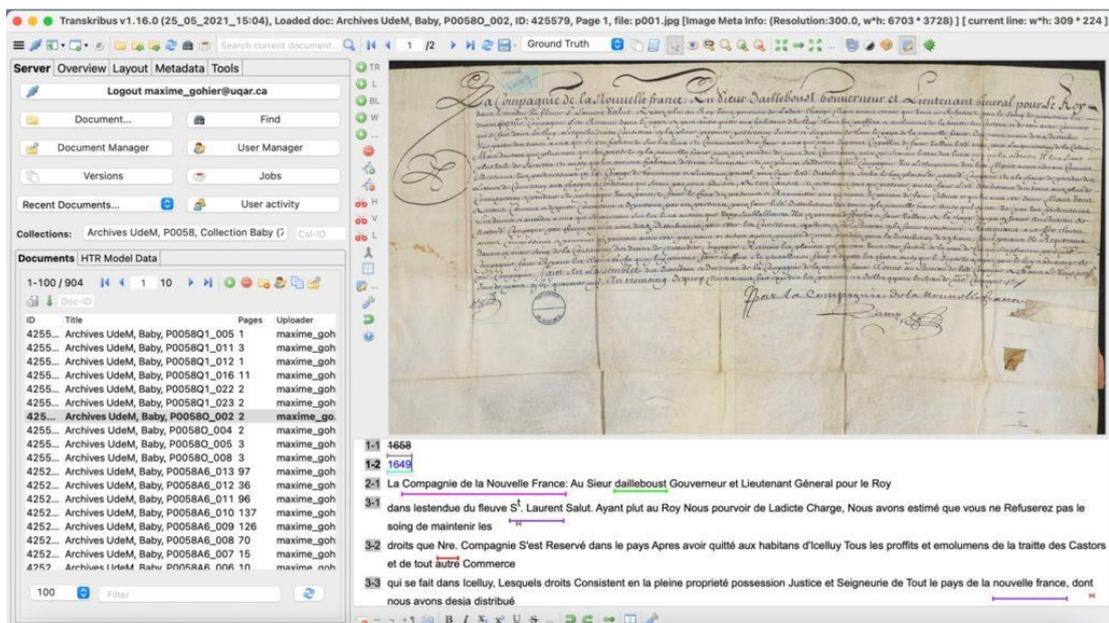


Figura 9: Interfaccia di Transkribus

¹⁴ <https://bv.informatik.uni-hamburg.de/amap#/>

Transkribus¹⁵ (Muehlberger 2019) è scaturito da due progetti europei (Transkriptorium, READ) ed è finanziato oggi da una società cooperativa e dagli utenti attraverso l'acquisto di crediti. Mette a disposizione previa registrazione e download del programma (TranskribusX) o via web (Transkribus Lite) strumenti per l'analisi del layout e del riconoscimento di testo con due engines (Pylaia basato su PyTorch e HTR+ basato su TensorFlow). Ci sono modelli allenati condivisi da altri utenti, ma è anche possibile allenare, previa richiesta, dei modelli propri, sia per layout specifici (P2PaIA¹⁶) sia per scritture particolari. Dispone inoltre dei tool per la ricerca FullText e Keyword e per l'importazione di grande quantità di trascrizioni pre-esistenti (Text2Image¹⁷). La piattaforma è parzialmente open source e l'utente può esportare i suoi documenti trascritti automaticamente in diversi formati, ma non il modello allenato. La società offre anche altri strumenti e servizi come lo ScanTent che permette di fotografare con lo smartphone documenti e trasferirli mediante app direttamente alla piattaforma, o l'allestimento di pagine web per la ricerca mediante keyword per progetti grandi (read&search).

Il più recente e promettente sviluppo nel campo è eScriptorium (in precedenza eScripta, Kiessling 2019), una piattaforma interamente OpenSource che ha avuto il primo collaudo proprio con scritture destra-sinistra come l'arabo e l'ebraico ed è dedicata espressamente al multigrafismo e multilinguismo. Similmente a Transkribus essa intende offrire una intera pipeline per il lavoro e incorporerà vari moduli già sviluppati in precedenza, attualmente Kraken¹⁸ per il riconoscimento di layout e di testo e in futuro Archetype (in precedenza The DigiPal Framwork)¹⁹ per annotazioni strutturate, utili per analisi paleografiche. Inoltre avrà strumenti per il post-processing come Pyrrha²⁰ per la post-correzione linguistica di corpora e TEI-Publisher²¹ per la pubblicazione dei risultati. È possibile l'import e l'esportazione anche di modelli. Allo stato attuale un accesso per utente singolo non è disponibile, ma la piattaforma richiede un'implementazione istituzionale.

¹⁵ <https://readcoop.eu/de/transkribus/>

¹⁶ <https://readcoop.eu/transkribus/docu/p2pala/>

¹⁷ <https://readcoop.eu/transkribus/docu/text2image/>

¹⁸ <https://github.com/mittagessen/kraken>

¹⁹ <https://github.com/kcl-ddh/digipal>

²⁰ <https://opensourcelibs.com/lib/pyrrha>

²¹ <https://teipublisher.com/index.html>

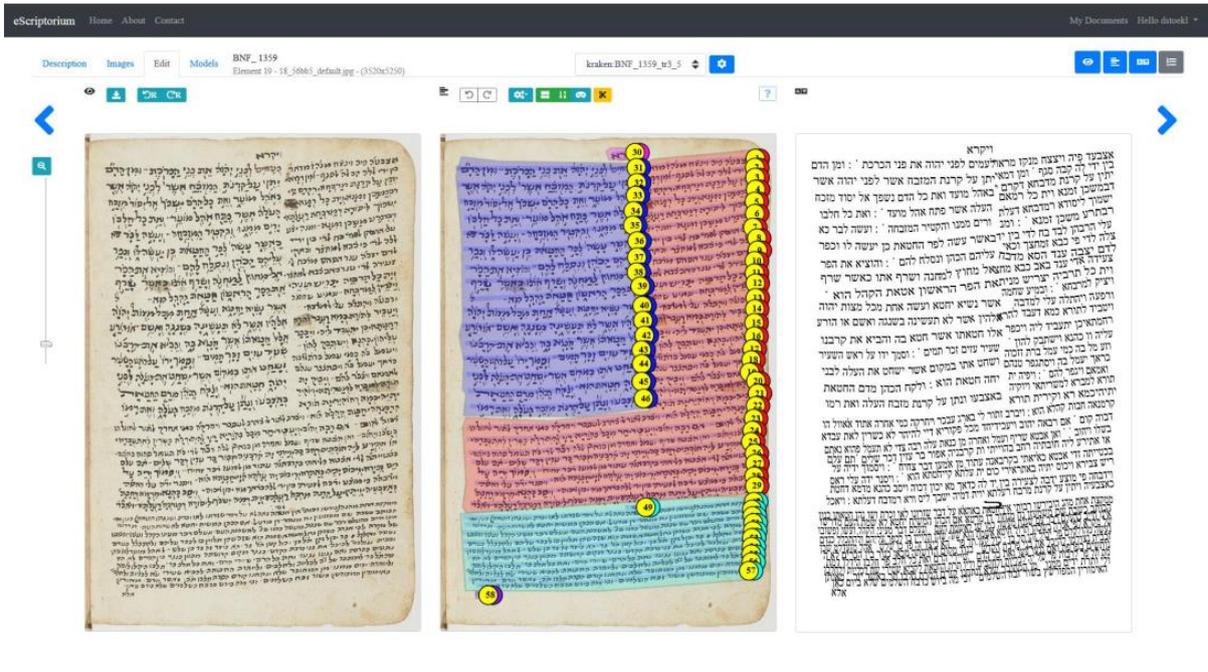


Figura 10: Interfaccia di eScriptorium

7. Riferimenti bibliografici

- Firmani 2018 = Donatella Firmani, Marco Maiorino, Paolo Meriardo, [et al.], *Towards Knowledge Discovery from the Vatican Secret Archives. In Codice Ratio -- Episode 1: Machine Transcription of the Manuscripts*, in *KDD '18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, New York, Association for Computing Machinery, 2018, p. 263–272. (<https://doi.org/10.1145/3219819.3219879>)
- Gabay 2021 = Simon Gabay, Jean-Baptiste Camps, Ariane Pinche, [et al.], *SegmOnto: common vocabulary and practices for analysing the layout of manuscripts (and more)*, in *1st International Workshop on Computational Paleography (IWCP@ICDAR 2021)*, Lausanne, 2021.
- Grieggs 2021 = Samuel Grieggs, Bingyu Shen, Greta Rauch, [et al.], *Measuring Human Perception to Improve Handwritten Document Transcription*, «IEEE Transactions on Pattern Analysis and Machine Intelligence», (2021), p. 1–1. (<https://doi.org/10.1109/TPAMI.2021.3092688>)
- Hodel 2021 = Tobias Hodel, David Schoch, Christa Schneider, [et al.], *General Models for Handwritten Text Recognition: Feasibility and State-of-the Art. German Kurrent as an Example*, «Journal of Open Humanities Data», 7 (2021), p. 13. (<https://doi.org/10.5334/johd.46>)
- Journet 2017 = Nicholas Journet, Muriel Visani, Boris Mansencal, [et al.], *DocCreator: A New Software for Creating Synthetic Ground-Truthed Document Images*, «Journal of Imaging», 3 (2017), 4, p. 62. (<https://doi.org/10.3390/jimaging3040062>)
- Lombardi 2020 = Francesco Lombardi, Simone Marinai, *Deep Learning for Historical Document Analysis and Recognition. A Survey*, «Journal of Imaging», 6 (2020), 10, p. 110. (<https://doi.org/10.3390/jimaging6100110>)
- Muehlberger 2019 = Guenter Muehlberger, Louise Seaward, Melissa Terras, [et al.], *Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study*, «Journal of Documentation», 75 (2019), 5, p. 954–976. (<https://doi.org/10.1108/JD-07-2018-0114>)

- Parziale 2020 = Antonio Parziale, Giuliana Capriolo, Angelo Marcelli, *One Step Is Not Enough: A Multi-Step Procedure for Building the Training Set of a Query by String Keyword Spotting System to Assist the Transcription of Historical Document*, «Journal of Imaging», 6 (2020), 10, p. 109. (<https://doi.org/10.3390/jimaging6100109>)
- Philips 2020 = James Philips, Nasseh Tabrizi, *Historical Document Processing: A Survey of Techniques, Tools, and Trends*., in *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, Budapest, Hungary, SCITEPRESS, 2020, p. 341–349. (<https://doi.org/10.5220/0010177403410349>)
- Riley 2021 = Chris Riley, UCL-University of Toronto: Transkribus, HTR, and medieval Latin abbreviations, Transcribe Bentham, <https://blogs.ucl.ac.uk/transcribe-bentham/2021/04/20/ucl-university-of-toronto-transkribus-htr-and-medieval-latin-abbreviations/>.
- Rajan 2019 = Vinodh Rajan, H. Siegfried Stiehl, *Making DIA Accessible to Non-Experts: Designing a Visual Programming Language for Document Image Analysis*, in *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, Sydney, Australia, IEEE, 2019, p. 23–27. (<https://doi.org/10.1109/ICDARW.2019.20048>)
- Kiessling 2019 = Benjamin Kiessling, Robin Tissot, Peter Stokes, [et al.], *eScriptorium: An Open Source Platform for Historical Document Analysis*, in *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, 2019, p. 19–19. (<https://doi.org/10.1109/ICDARW.2019.10032>)

8. Elenco delle figure

Figura 1: Modello di rete neurale

MultiLayerNeuralNetwork_english.png: Chrislb derivative work: — HELKNOWZ | TALK | enWP TALK, CC BY-SA 3.0 <<https://creativecommons.org/licenses/by-sa/3.0/>>, via Wikimedia Commons: https://commons.wikimedia.org/wiki/File:MultiLayerNeuralNetworkBigger_english.png

Figura 2: Tipica architettura di una rete neurale convoluzionale

Aphex34, CC BY-SA 4.0 <<https://creativecommons.org/licenses/by-sa/4.0/>>, via Wikimedia Commons: https://commons.wikimedia.org/wiki/File:Typical_cnn.png

Figura 3: Possibile workflow per il processing di documenti storici

Elaborazione propria sulla base di Philips 2020, p. 3.

Figura 4: Linea di base

Max Naylor, Public domain, via Wikimedia Commons: https://commons.wikimedia.org/wiki/File:Typography_Line_Terms.svg

Figura 5: Iniziale dal Cathach di San Columba (sec. VIII)

Cathach of St Columba, Public domain, via Wikimedia Commons: <https://commons.wikimedia.org/wiki/File:CathachOfStColumba.jpg>

Figura 6: Keyword spotting

Felix Dietrich, Search collections with Transkribus, in occasione della Transkribus User Conference 2020, 6 febbraio 2020, <https://www.youtube.com/watch?v=42FiXv37Ggl>

Figura 7: Esempio di keyword spotting in HIMANIS

<http://himanis.huma-num.fr/app//index.php/ui/show/chancery/69/155?q=notari%2A&t=50>

Figura 8: Interfaccia di AMAP

Rajan 2019 = Vinodh Rajan, H. Siegfried Stiehl, *AMAP: A Visual Programming Language Based System to Support Document Image Analysis*, in *Proceedings of Mensch und Computer 2019*, New York, Association for Computing Machinery, 2019, p. 881–884, fig. 1.

(<https://doi.org/10.1145/3340764.3345372>)

Figura 8: Interfaccia di Transkribus

<https://readcoop.eu/success-stories/nouvelle-france-numerique-collaboration-and-partnership-arising-from-ai/>

Figura 10: Interfaccia di eScriptorium

Daniel Stoekl Ben Ezra, Peter Stokes, Marc Bui, [et al.], eScriptorium: An Open Source Platform for Historical Document Analysis, 30 marzo 2021,

<https://www.slideshare.net/biblissima/escriptorium-an-open-source-platform-for-historical-document-analysis>