



HAL
open science

Humanities Computing and Legal Informatics in Italy

Deborah Grbac

► **To cite this version:**

Deborah Grbac. Humanities Computing and Legal Informatics in Italy. Whitney Kramer, Iliana Burgos, and Evan Muzzall. Text and Data Mining Literacy for Librarians, ACRL, Association of College and Research Libraries, In press, 9798892555951. <hal-05229543>

HAL Id: hal-05229543

<https://hal.science/hal-05229543v1>

Submitted on 29 Aug 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Humanities Computing, Legal Informatics, and Text Analysis Pedagogy in Italy: A Brief History of the Practice

Deborah Grbac

Text analysis in Italy developed from practical experiences that gradually introduced new techniques, instruments, and cooperative networks in different fields: computational linguistics (through the development of the Italian Digital Humanities and Digital Culture Association AIUCD), governmental (the governmental digital platforms giving access to legal texts—case laws *Italgiure* and laws *Normattiva*), and academic, with the spreading of different research centers working in humanities computing. This chapter discusses the evolution of humanities computing and legal informatics in Italy as well as some practical examples of how the library at the Catholic University of the Sacred Heart in Milan (Università Cattolica del Sacro Cuore di Milano) has engaged text analysis pedagogy over the last eight years.

Humanities Computing and Legal Informatics in Italy

In 1949, linguist Father Roberto Busa S. J. led the research project, *Index Thomisticus*, to create an index from the writing of Saint Thomas. Through this project, Busa was one of the first to use IBM punch card machines to analyze texts.¹ This project was foundational for the field of humanities computing in Italy and catalyzed the establishment of research centers engaging automation and information retrieval. Circa 1956, Father Busa founded the Center for Automation of Literary Analysis (*Centro per l'automazione dell'analisi letteraria*) in Gallarate, Italy.² In the 1950s and 1960s, the center was active in projects providing machine translation and the information retrieval protocol called KWIC, or keywords-in-context.³

After 1978, Father Busa presented workshops at the Group of Interdisciplinary Research for the Computational Study of Words and Expressions (*Gruppo interdisciplinare di ricerche per la computerizzazione dei segni e dell'espressione*, GIRCSE) at the Catholic University of the Sacred Heart in Milan. The Center for Automation of Literary Analysis was active until the mid-eighties and one of the first research groups focused on computational linguistics in Italy, GIRCSE, became the Center for Interdisciplinary Research on the Computational Study of Words and Expressions (*Centro interdisciplinare di ricerche per la computerizzazione dei segni e dell'espressione*), one of the research centers at the Catholic University of the Sacred Heart in Milan.

Computational linguistics became increasingly prevalent in research studies in Italy from the 1970s through the 1990s. During this time, several computational linguistics institutes were created. Computational Linguistics Institute (*Istituto di Linguistica Computazionale*, ILC) was established in Pisa and led by Antonio Zampolli, Father Busa's pupil. The "Roman school" was established at La Sapienza University and led by Fabio Ciotti. In 1980, the ILC became the European Intellectual Lexicon Institute (*Istituto per il lessico intellettuale europeo*), an institution that is part of the Italian National Research Center (*Consiglio Nazionale Ricerche*, CNR).

In the following years, new schools and institutions focused on both computational linguistics and quantitative methods for textual analysis. For instance, in 2007, Padua University established a center with an interdisciplinary approach, entitled Interdisciplinary Group in Text Analysis (*Gruppo Interdisciplinare di Analisi Testuale*, GIAT). The center is associated with the multidisciplinary Department of Philosophy, Sociology, Pedagogy and Applied Psychology (*Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia Applicata*) and has organized the Summer School in Quantitative Analysis of Textual Data since 2013.

The Italian Association of Digital Humanities and Digital Culture (AIUCD) was created in 2011.⁴ Under the direction of Professor Fabio Ciotti, the association has published *Umanistica Digitale*, a scientific journal focused on digital humanities discourse, since 2017.⁵

Legal Informatics

In 1962, the Center for Automatic Documentation (*Centro per la documentazione automatica*) in Milan hosted the first documented instance of performing computational

analysis on legal documents in Italy. Different projects followed, each aimed at working on automatic analysis of legal cases. In the 1970s, a database called Italgire⁶ was created by the Italian Ministry of Justice. It is currently managed by the Italian Supreme Court, Electronic Documentation Center.⁷

At the same time, the CNR created two departments, an Automatic Department (later called Legal Informatics), and a Legal Lexicographic Department, called Legal Documentation Institute (Istituto per la Documentazione Giuridica, IDG). The latter has been active for the first fifty years of Legal Informatics in Italy. Today, it is called Legal Informatics and Legal Systems Institute (Istituto di Informatica Giuridica e Sistemi Giudiziari, IGSG), CNR, Florence.⁸

The IGSG is also in charge of the bibliographic database referencing all Italian periodicals on law and case law, the DO.GI., or Legal Doctrine, scholarly publications on law.⁹

The IDG, Legal Documentation Institute, began publishing the first scholarly journal on legal informatics published in Italy, now called *Rivista italiana di informatica e diritto* (*Italian Journal of Informatics and Law*).¹⁰ In 2022, the IGSG published a monograph: *La Internet Governance e le sfide della trasformazione digitale*, in cooperation with the Internet Society, Italy Chapter on Internet Governance. It gives a state-of-the-art discussion of internet development under different legal aspects in Italy and the world.¹¹

Since 2010, the Italian Government has been publishing online Italian law on the Normattiva platform. Normattiva is a textual database containing all the in-course and former Italian laws, managed by Polygraph Institute and State Mint.¹²

Software

The first two Italian-language software tools for text analysis were TalTaC (Trattamento automatico lessicale e testuale per l'analisi del contenuto),¹³ developed by Professor Sergio Bolasco from Sapienza University in Rome in 1999, and T-LAB launched by Franco Lancia from Frosinone in 2001.¹⁴ Both packages use linguistic, textual, and statistical tools to analyze texts, following two approaches: a text analysis (lexical and textual metrics) and a text-mining approach (information retrieval and information extraction). They are aimed at “representing” texts both at the word level (tokens) and context level (part of document, a full document, or sub-corpora) by offering information about language, key terms used, and the contents.¹⁵

For nearly twenty years (2000–2019), TalTaC, under the direction of Professor Sergio Bolasco, offered instruction on software and practical text analysis issues at the Sapienza University in Rome. Since its creation, T-LAB has been presented and used in courses held in laboratories at Sociology and Psychology Departments in different universities.

Today, text analysis workflows are more customizable than ever. Thanks to an increase in resources on text analysis and a growing number of tools, it is possible to retrieve, classify, summarize, and represent content in a way that efficiently identifies useful documents and analyzes them. This enables researchers to design research workflows suitable for specific investigations.

Early Examples of TDM Tools for Information Retrieval and Information Extraction

ITHAKA JSTOR¹⁶ was one of the first library vendors to use text analysis methods to showcase content visualizations to users. It introduced the Text Analyzer tool in 2017 (see figure 10.1). The tool is still active as of the writing of this chapter and allows users to upload a full-text document (such as a Word document or PDF) in different languages to query the JSTOR Archive for similar content.¹⁷

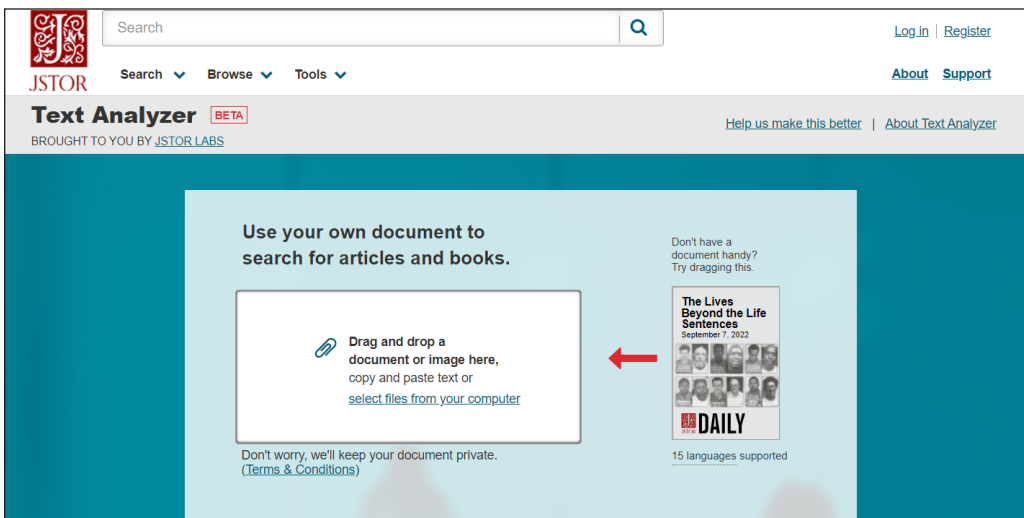


Figure 10.1. The ITHAKA JSTOR Text Analyzer interface.¹⁸

The tool’s algorithm processes both the texts uploaded by the user to build a query and the content of the database in order to find suitable matches. The tool is an example of merging information retrieval and information extraction processes. The tool extracts information from a source document and then suggests further readings based on the content of the source document.

As part of the search results preview, Text Analyzer automatically conducts topic modeling and Named Entity Recognition (NER)¹⁹ on the input document. The five most significant topics and named entities are used to perform a query²⁰ on JSTOR Archive documents, enabling users to distant read them and find related texts. The results are visualized by the Equalizer, a fine-tuning tool enabling further adjustment of the results according to the topics or named entities identified.

The user can weigh the list of controlled terms or “topics” proposed by the tool to further their search. The final results are a list of reading suggestions derived from the topics and entities extracted from the uploaded text, scored by the weight of the term in the input text and the importance of the term in the retrieved document, and results are proposed according to the co-occurrence of terms between the input text and the recommended results.²¹

Another example of an early text analysis application used for information extraction is the LinkedSDG service, which uses linked data derived from the United Nations' Sustainable Development Goals (SDGs).²² It was developed in 2019 by DESA, the Department of Economic and Social Affairs of the United Nations, in cooperation with the United Nations Dag Hammarskjöld Library in New York.²³ A user uploads a document in the LinkedSDG app and then retrieves a calculation of the percentage of SDGs present in the text. The tool is used to verify whether the uploaded documents contain terms that correlate with metadata of the SDGs and measures the presence of SDG-related terms. Additional visualizations propose NER (geographical names) and links to concepts related to Sustainable Development Goals derived by way of an ad hoc taxonomy. The tool also proposes a list of statistical datasets related to SDGs retrieved by human-curated semantic annotations. Both the United Nations Library Thesaurus (UNBIS Thesaurus) and the SDGs taxonomy are published and maintained as linked open data.²⁴

In 2019, the GoTriple project launched the European GoTriple aggregation platform, a discovery service dedicated to social sciences and humanities for information retrieval and information extraction. The service allows users to navigate and browse different types of research products in twelve European languages, each offering different services for social sciences and humanities research. These research products include literature, datasets, research projects, and researchers' profiles within the platform. With a single query, users can retrieve results from a collection of open access scholarly communications products in eleven European languages. Users can also access tools including search results visualizations, annotations, and recommendations. The information is extracted by a topic-modeling classifier ("Triple topics"), that uses the Triple Vocabulary for annotations.²⁵ Results are retrieved by means of the co-occurrence between keywords used in the query and words present in the retrieved documents. The results are rendered by visualizations, including a knowledge map²⁶ that is represented by bubbles of documents aggregated by common words—a streamgraph,²⁷ a chronological presentation of the evolution of the keywords inside the texts, plotting the number of documents present in the repository per keyword (see figure 10.2).

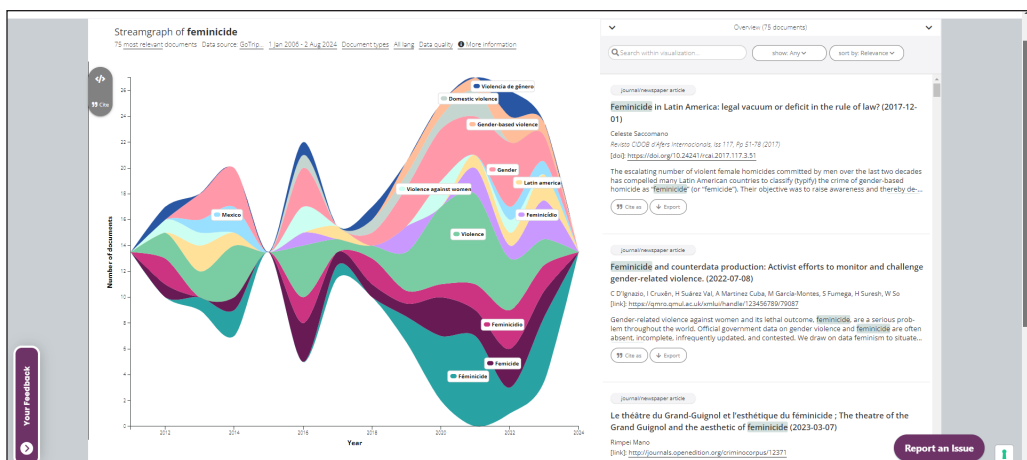


Figure 10.2. A GoTriple Streamgraph representation, with an example proposed on the website, https://www.gotriple.eu/streamgraph?q=feminicide&from=2006-01-01&to=2024-08-02&lang_id=all-lang, last accessed 2024/08/23rd.

Teaching Text Analysis

Since 2012, I have taught information literacy to the Political and Social Sciences School at Catholic University of the Sacred Heart in Milan. I use practical examples to explain to students that once they know how information is organized, they may retrieve it in different ways, not only by indexes but also by using unstructured text data.

My initial approach was to gradually introduce students to the broader concept of data. I began with the library catalog as a database and used metadata as an example of data. After explaining how to search by means of indexes in the Classic Catalog, I explained to students how to use keyword search to extract information from metadata, be imaginative, or exploit all the opportunities given by a keyword search on a bibliographic record.

I shifted my approach in 2017 after attending the IQLA-GIAT Summer School in Quantitative Analysis of Textual Data at Padua University.

The Summer School deals with quantitative linguistic methods and textual data analysis and each year focuses on a different methodological focus. In 2017, the focus was topic detection and authorship attribution. After this experience, I started to approach text analysis tools and see text analysis applications under the hood of tools at my disposal. For example, I was able to notice and fully grasp the text analysis methods featured in JSTOR's Text Analyzer tool.

In spring 2020, JSTOR Labs asked me to participate as a beta-evaluator in an experimental series of text analysis classes to help them build an educational service for TDM. I accepted this invitation and actively participated in the partnership that would launch Constellate, JSTOR's cloud-based environment for learning and teaching text analysis. Constellate enables users to store, reuse, and share Jupyter Notebooks, which contain Python code to help perform different text analysis tasks.²⁸ The collaboration enabled me to grow my skills from performing some analysis automatically using third-party tools to developing an understanding of what kinds of analyses I was performing and why.²⁹

The exercise of learning by doing helped me shift my courses from information literacy to data literacy and from information retrieval to text analysis in 2023. It was no more a matter of simply retrieving metadata using information extraction or considering only metadata as data, but I was able to introduce my students to the ideas of textual data, data science, and text analysis as research methods.

In October 2023, I took a significant step in my professional growth when I addressed computational linguistics students. This experience highlighted a paradigm shift in my role: viewing text as data means that information retrieval is no longer solely a librarian's purview but also a natural language processing (NLP) task. Consequently, the librarian's role must evolve beyond traditional assignment indexing and metadata retrieval to actively engage emerging TDM practices and purposes.

Conclusions

My experiences in TDM have been possible through the theory I learned at the IQLA-GIAT Summer School in Quantitative Analysis of Textual Data and the practice I developed

cooperating in different projects. Serving as a beta-evaluator with JSTOR Labs, acting as a depository librarian for the United Nations Depository Library System, and beta-testing the GoTriple project gave me opportunities to explore and understand TDM tools from the perspectives of the developers and creators. However, even if theory was essential, hands-on practice enabled me to understand TDM more deeply. As I built knowledge in TDM, I was constrained by the distinct separation of text analysis into different epistemic communities in Italy. Disciplinary communities of linguists collaborate with other linguists and occasionally with digital humanists. Textual statisticians often work with sociologists or economists. Legal informaticians primarily work with administrative institutions. Rarely do these communities collaborate. I have had the opportunity to speak at conferences across all three of these scientific communities and contribute to various scientific journals, yet I consistently observe their seclusion.

Similarly, I observe that the community of Italian librarians is also isolated from other professions. Text analysis is not yet widespread in our field. There is a trend in Italian librarianship to prioritize bibliometric analysis, due to annual university research evaluation campaigns, which emphasize citation metrics to indicate scholarly impact. Currently, this community is cautiously exploring artificial intelligence (AI) applications, often with skepticism and a lack of foundational knowledge in text analysis. However, I argue that text analysis literacy is crucial for interpreting AI applications like LLMs (large language models) or generative AI.

At the March 2024 Stelling Biblioteche Conference, only one paper at the proceedings engaged with text data literacy or data literacy more broadly. It seems that my colleagues across academic libraries in Italy are just beginning to engage the notion that librarians can be educators of information literacy as well as data and other digital literacies.³⁰ Having benefited from international text analysis communities of practice, I understand the need to share text analysis knowledge across the profession. Librarians based in Italy can work collaboratively to support the text analysis work already underway across our institutions.

Acknowledgments

My thanks go to Irene Saonara, PhD, who, with her PhD thesis in 2014, first introduced me to text analysis, and Professor Arjuna Tuzzi, Università di Padova, for accepting me in 2017 at the IQLA-GIAT International Summer School in Quantitative Analysis of Textual Data. Professor Maria Spano, Università di Napoli Federico II, Professor Michelangelo Misuraca, Università della Calabria, Professor Giuseppe Giordano, Università di Salerno, for accepting me at JADT 2022 Conference. JSTOR Labs at ITHAKA, and particularly Amy J. Kirchoff, Nathan Kelber, Zhuo Chen, and Ken DiFiore. For the INCONECSS (International Conference on Economics and Business Information) community, Tamara Pianos and Thorsten Meyer, and other former NEREUS (European Network of Social Sciences Libraries) colleagues. For the United Nations Depository Libraries System community and United Nations Library and Archives in Geneva: Armando Da Silva, Cristina Giordano, Hermine Diebolt, and Ramona Kohrs.

Notes

1. Tito Orlandi and Francesca Tomasi, “Una storia dell’informatica umanistica in Italia,” in *Digital Humanities. Metodi, strumenti, saperi*, Collana Studi superiori, n. 1376, Teoria della letteratura e critica letteraria, ed. Fabio Ciotti (Roma: Carocci editore, 2023), 35–37, 40; Marco Passarotti and Julianne Nyham, eds., *One Origin of Digital Humanities: Fr. Roberto Busa in his own words* (Cham: Springer International Publishing, 2019), XIX–XX. In 2010, Father Busa donated his library and his working papers to Università Cattolica del Sacro Cuore in Milan: the collection today provides an extremely relevant source for reconstructing both the scholar’s activity and the birth of applying computer science to linguistics. Inside the personal archive one can find very rich documentation consisting of thousands of documents spanning through the second half of the last century.
2. Passarotti and Nyham, eds., *One Origin of Digital Humanities*, 69.
3. KWIC is the “Key Word In Context Protocol” created by Peter Luhn at IBM in cooperation with Father Busa; Passarotti and Nyham, eds., *One Origin of Digital Humanities*; nowadays, the KWIC protocol is commonly used for finding out collocations of words in a text or collection of texts.
4. AIUCD, Associazione per l’Informatica Umanistica e la Cultura Digitale, accessed December 12, 2021, <http://www.aiucd.it/>.
5. Orlandi and Tomasi, “Una storia dell’informatica umanistica in Italia,” 46.
6. “Italgire” is a compound word that synthesizes the words in the phrase “la giurisprudenza italiana” (Italian case laws).
7. “Corte Di Cassazione,” ItalgireWeb, accessed December 21, 2023, https://www.italgiure.giustizia.it/index_en.asp?lang=en.
8. Ginevra Peruginelli and Mario Ragona, “Forward” in English, in Ginevra Peruginelli and Mario Ragona, eds., *L’Informatica giuridica in Italia. Cinquant’anni di studi, ricerche ed esperienze*, Collana dell’Istituto di Teoria e Tecniche dell’Informazione Giuridica del Consiglio Nazionale delle Ricerche, Serie “Studi e documenti”, n. 12 (Napoli: Edizioni Scientifiche Italiane, 2014), 29–32.
9. The DO.GI. database, Italian juridical periodicals bibliographical database, accessed August 8, 2023, <http://dati.igsg.cnr.it/dogi>.
10. It was of the first directors of the IDG, Professor Luigi Lombardi Vallauri, the introduction already in the seventies of the world “algorithm” in the legal information field in an article published in the journal “Informatica e diritto,” *Cfr, supra*, 13–24.
11. Laura Abba, Adriana Lazzaroni, and Marina Pietrangelo, *La Internet Governance e le sfide della trasformazione digitale* (Napoli: Editoriale Scientifica, 2022).
12. Istituto Poligrafico e Zecca dello Stato oversees law publishing and money coinage; Normattiva, accessed December 12, 2023, <https://www.normattiva.it/>.
13. TalTaC, Natural Language and Textual Processing for Contents Analysis, accessed August 8, 2024, <https://www.taltac.com/presentazione/>.
14. T-LAB, accessed August 8, 2024, <https://www.tlab.it/?lang=eng>.
15. The resulting analyses are content representations overlapping different approaches and tools used. Analysis results are proposed as feature matrices that can be further analyzed statistically, and outputs can be visualized using different visualizations, including graphs, dendrograms, world clouds, and aerogrammes.
16. JSTOR is an early content provider of both the metadata about and full texts of scholarly communications. JSTOR, accessed August 23, 2024, <https://www.jstor.org/>.
17. ITHAKA graduated the “Text Analyzer” on August 30, 2024. JSTOR, accessed April 24, 2025, <https://labs.jstor.org/projects/text-analyzer/>.
18. “JSTOR Labs Text Analyzer,” JSTOR, accessed August 23, 2024. <https://www.jstor.org/analyze/>.
19. Topic modeling is a text mining technique using different methods to uncover contents into texts, both explicitly, when topics are identified by means of a thesaurus and annotations, or inferred when algorithms are used to extract contents, as for instance the Latent Dirichlet Allocation, an unsupervised model. Named entities are names identifying persons, organizations, geographical locations, even dates, they are identified into text using services or tools extracting this kind of information.

20. In the Text Analyzer, the TF-IDF calculation is used to choose the NER, the acronym TF-IDF stands for Term-Frequency against Inverse-Document Frequency, or the weight of a word against all the weights of the words in the documents considered.
21. Ron Snyder, *Under the Hood of Text Analyzer*, March 7, 2017, JSTOR Labs, <https://labs.jstor.org/blog/under-the-hood-of-text-analyzer-2/>, accessed August 23, 2024.
22. Sustainable Development Goals are seventeen goals defined in the “2030 Agenda for Sustainable Development” adopted in 2015 by United Nations, with the aim of contributing to peace and prosperity for people and the planet by means of an urgent call of action on issues as for instance: eradicating poverty and hunger, improving health and education; “The 17 Goals,” United Nations, DESA, <https://sdgs.un.org/goals>, accessed August 23, 2024.
23. LinkedSDG, <https://linkedsdg.officialstatistics.org/#/>; for more information about the service, <https://sustainabledevelopment.un.org/LinkedSDGs/about>, accessed August 23, 2024.
24. The UNBIS Thesaurus is a database of the controlled vocabulary commonly used to describe United Nations documents; it is expressed in six official languages and it develops continuously according to the introduction of new topics in the United Nations activities agenda. “UNBIS Thesaurus and Linked Data Service,” United Nations Dag Hammarskjöld Library, accessed August 23, 2024, <https://research.un.org/en/thesaurus>, accessed August 23, 2024.
25. The Triple Vocabulary is a multilingual set of concepts consisting in a subset of the Library of Congress Subject Headings, <https://www.semantics.gr/authorities/vocabularies/SSH-LCSH/?language=en>.
26. Open Knowledge Map, <https://openknowledgemaps.org/faq>.
27. GoTriple, <https://gotriple.eu/>; for further information on the working of the discovery platform, “About TRIPLE,” GoTriple, <https://gotriple.eu/about>, accessed August 23, 2024.
28. Constellate Homepage, <https://constellate.org/>; for further Constellate applications, refer to chapter 29 written by Jajwalya Karajgikar, “Constellate Support for Multilingualism” in this book.
29. ITHAKA sunsetted “Constellate” on July 1, 2025. Constellate, “Constellate Sunset,” <https://constellate.org/docs/constellate-sunset?ref=cms-prod.constellate.org>, accessed April 24, 2025.
30. Elena Almangano, Eleonora De Longis, and Damiano Orru, “Dalla data literacy all’AI literacy: bibliotecari educatori e formatori,” in *Biblioteche e nuove forme della lettura. Dal ruolo strategico della literacy alle trasformazioni in atto nell’ecosistema digitale. Relazioni di mercoledì 20 e giovedì 21 marzo*, Convegno delle Stelline 2024 (Milano: Associazione Biblioteche oggi, Editrice Bibliografica, 2024), 131–39.

Bibliography

- Abba, Laura, Lazzaroni Adriana, and Marina Pietrangelo. *La Internet Governance e le sfide della trasformazione digitale*. Napoli: Editoriale Scientifica, 2022.
- AIUCD. <http://www.aiucd.it/>, last accessed 2024/08/23rd.
- Bauder, Julia, ed. *Data Literacy in Academic Libraries. Teaching Critical Thinking with Numbers*. Chicago: ALA (American Library Association) Editions, 2021.
- Bolasco, Sergio. *L’analisi automatica dei testi. Fare ricerca con il text mining*. Collana Studi Superiori, 922, Statistica. Roma: Carocci editore, 2013.
- Ciotti, Fabio, ed. *Digital Humanities. Metodi, strumenti, saperi*. Collana Studi superiori, n. 1376, Teoria della letteratura e critica letteraria. Roma: Carocci editore, 2023.
- Constellate. <https://constellate.org/>.
- Convegno delle Stelline 2024. *Biblioteche e nuove forme della lettura. Dal ruolo strategico della literacy alle trasformazioni in atto nell’ecosistema digitale. Relazioni di mercoledì 20 e giovedì 21 marzo*. Milano: Associazione Biblioteche oggi, Editrice Bibliografica, 2024.
- DO.GI. <http://dati.igsg.cnr.it/dogi>.
- GIAT. <https://www.fisppa.unipd.it/giat-it/home>.
- GoTriple. <https://gotriple.eu/>.
- Grimmer, Justin, Margaret E. Roberts, Brandon M. Steward. *Text as Data. A New Framework for Machine Learning and the Social Sciences*. Princeton: Princeton University Press, 2022.
- ItalggiureWeb. “Corte Di Cassazione.” Accessed December 21, 2023. https://www.italgiure.giustizia.it/index_en.asp?lang=en.

- JSTOR. Accessed August 8, 2024. <https://www.jstor.org/>.
- JSTOR Labs. “Text Analyzer.” Accessed August 23, 2024. <https://www.jstor.org/analyze/>.
- LaFleur, Marcelo T. *Art is Long, Life is Short: An SDG Classification System for DESA Publications*. DESA Working Paper No. 159, United Nations Symbol ST/ESA/2019/DWP/159, 2019.
- Lancaster, Wilfrid F. *Indexing and Abstracting in Theory and Practice*. 3rd ed., London: Facet Publishing, 2003.
- LinkedSDG. <https://linkedsdg.officialstatistics.org/#/>.
- Normattiva. <https://www.normattiva.it/>.
- Open Knowledge Map. <https://openknowledgemaps.org/faq>.
- Passarotti, Marco, and Julianne Nyham, eds. *One Origin of Digital Humanities: Fr. Roberto Busa in His Own Words*. Cham: Springer International Publishing, 2019.
- Peruginelli, Ginevra, and Mario Ragona, eds. *L’Informatica giuridica in Italia. Cinquant’anni di studi, ricerche ed esperienze*. Napoli: Edizioni Scientifiche Italiane, 2014.
- Snyder, Ron. *Under the Hood of Text Analyzer*. March 7, 2017. <https://labs.jstor.org/blog/under-the-hood-of-text-analyzer-2/>.
- Sprugnoli, Rachele, and Ježek Elisabetta. *Linguistica computazionale. Introduzione all’analisi automatica dei testi*. Bologna: Il Mulino, 2023.
- TalTaC. <https://www.taltac.com/presentazione/>.
- T-LAB. <https://www.tlab.it/?lang=eng>.
- Tuzzi, Arjuna. *L’analisi del contenuto: introduzione ai metodi e alle tecniche di ricerca*. Roma: Carocci, 2003.
- United Nations Dag Hammarskjöld Library. “UNBIS Thesaurus and Linked Data Service.” Accessed August 23, 2024. <https://research.un.org/en/thesaurus>.
- United Nations, DESA. “The 17 Goals.” <https://sdgs.un.org/goals>.