

Per un'analisi della rivista *Umanistica Digitale* in ottica DEI

Rossana Morriello¹, Lucia Sardo²,

¹ Università degli Studi di Firenze, Italia – rossana.morriello@unifi.it

² Alma Mater Studiorum Università di Bologna, Italia – lucia.sardo@unibo.it

ABSTRACT (ITALIANO)

L'intervento presenta il risultato di una analisi sugli articoli della rivista *Umanistica Digitale* dalla sua fondazione a oggi in un'ottica DEI, considerando che questo è il tema del convegno che sarà organizzato nel 2025. L'analisi ha rilevato una grande ricchezza di tematiche trattate ma una scarsa presenza di articoli dedicati alle tematiche DEI; inoltre, l'utilizzo di parole chiave non uniformate rende difficoltosa la ricerca di testi su argomenti simili e riduce la *discoverability* degli articoli. Il contributo propone l'uso di un vocabolario strutturato per la corretta disseminazione delle tematiche DEI.

Parole chiave: DEI, *Umanistica Digitale*, keywords, text mining

ABSTRACT (ENGLISH)

For an analysis of Umanistica Digitale journal from a DEI perspective

The paper presents the result of an analysis on the articles of the journal *Umanistica Digitale* from its foundation to the present day from a DEI perspective, considering that this is the theme of the conference to be organised in 2025. The analysis revealed a great richness of topics treated but a low presence of articles dedicated to DEI topics; furthermore, the use of non-uniform keywords makes it difficult to search for texts on similar topics and reduces the discoverability of the articles. The contribution proposes the use of a structured vocabulary for proper dissemination of DEI topics.

Keywords: DEI, *Umanistica Digitale*, keywords, text mining

1. INTRODUZIONE

Le keywords svolgono diverse funzioni importanti nell'era dei dati e delle applicazioni dell'intelligenza artificiale, tra cui innanzitutto permettere la *discoverability* e la *findability* delle risorse attraverso i motori di ricerca e nei sistemi di information retrieval delle banche dati. Sono quindi alla base delle strategie di Search Engine Optimization (SEO) per diverse tipologie di risorse. Nell'ambito della ricerca accademica e delle banche dati scientifiche le keywords favoriscono quindi il recupero, la disseminazione e la condivisione dei risultati che è il fine ultimo della ricerca scientifica. Consentono di mettere in contatto ricercatori che si occupano di tematiche affini poiché tramite le parole chiave si possono cercare articoli di argomento simile a ciò che è di interesse, ed eventualmente citarli, contribuendo alla determinazione degli indicatori bibliometrici. La presenza di keywords favorisce l'applicazione di strumenti di data analysis e information retrieval avanzati, sempre più basati sull'intelligenza artificiale.

Per queste ragioni gli editori nelle loro banche dati, siano esse citazionali, testuali, bibliografiche, attribuiscono grande importanza a questa forma di indicizzazione semantica che consente loro di organizzare efficacemente i contenuti e di raggrupparli in cluster per agevolare innumerevoli forme di data analysis, oltre alla *discoverability*, e di aggiornare costantemente l'organizzazione dei contenuti digitali sulla base dell'evoluzione della ricerca scientifica.¹ Gli articoli contenuti nelle principali banche dati scientifiche sono infatti dotati di varie keywords, o key phrases, come sarebbe più opportuno chiamarle poiché spesso si tratta di frasi composte da più parole (Hartley & Kostoff, 2003), insieme al titolo e all'abstract dell'articolo. In misura crescente, questi elementi paratestuali, e in particolare le key phrases, vengono utilizzate per l'assegnazione della peer review a determinati revisori, tramite il matching tra le keywords di un articolo e le keywords con le quali il revisore ha sintetizzato le sue competenze, come spesso avviene sulle piattaforme editoriali. Per motivi analoghi le keywords sono fondamentali nei progetti di ricerca che si sottopongono per il finanziamento e che richiedono revisione paritaria.

Per quanto riguarda il nostro macrotema, ovvero diversità, equità e inclusione (DEI) le keywords sono tra i metadati che consentono di ricondurre le pubblicazioni agli ambiti DEI, poiché permettono di costruire, insieme agli altri elementi paratestuali, il contesto e il significato necessari al lettore per accedere consapevolmente all'informazione su questi temi. Nell'acronimo DEI, con "diversità" si intende una serie di

¹ Per esempio, è quanto dichiara esplicitamente Elsevier nella banca dati SciVal, che offre soluzioni di research intelligence e data visualization sui dati di Scopus, a cui faremo riferimento più avanti in questo contributo.

differenze socio-economiche e culturali potenzialmente svantaggiose quali disabilità, età, genere, orientamento sessuale, status socio-economico, lingua, pratiche di alfabetizzazione, background educativo, localizzazione geografica, razza, etnia, cultura e religione (Raju 2020; Morriello & Sardo 2023), e l'“inclusione” insieme all'“equità” sono aspetti essenziali e integrali ai processi volti all'incremento della diversità nel mondo dell'informazione, della ricerca e dell'educazione (Raju 2020). Implicita nell'acronimo è l'azione svolta per superare le barriere e migliorare la diversità. Dunque, non si tratta solo di mappare il soddisfacimento dei criteri di diversità, equità inclusione in un settore quale può essere il settore o i settori disciplinari di copertura di una rivista scientifica, ma anche e soprattutto di agire per potenziare la diversità. Con tale obiettivo, costruttivo e non certamente critico, proponiamo il nostro contributo.

2. OBIETTIVI E METODOLOGIA

L'intervento presenta un'analisi degli articoli della rivista *Umanistica Digitale* dalla sua fondazione a oggi, per rilevare quanto e come le tematiche riconducibili a diversità, equità e inclusione siano presenti e in che modo vengano indicizzate tramite le keywords.

L'analisi è stata condotta utilizzando le informazioni e i dati presenti sul sito della rivista stessa. Nello specifico sono stati presi in considerazione gli autori e le parole chiave inserite negli articoli, al fine di rilevare le keywords legate ai temi DEI. Inoltre, sono stati rilevati i dati relativi al genere fra coloro che hanno presentato contributi alla rivista.

La raccolta ed elaborazione dei dati è stata effettuata utilizzando il software open source OpenRefine, con una successiva integrazione e verifica manuale nel caso in cui sia reso necessario. A questa fase è seguita l'analisi di quanto emerso dai dati raccolti, con una riflessione particolare sull'uso delle keywords. La sezione finale dell'intervento introduce una proposta per un ripensamento nell'uso delle parole chiave, al fine di facilitare il reperimento di documenti rilevanti e il posizionamento degli studiosi/e nei rispettivi ambiti disciplinari.

I dati analizzati con OpenRefine sono poi stati confrontati con i dati di Scopus e di SciVal. Da settembre 2021, infatti, la rivista *Umanistica Digitale* è indicizzata nella banca dati citazionale Scopus di Elsevier e di conseguenza nella banca dati SciVal, dello stesso editore, che attinge ai dati di Scopus per offrire soluzioni di research intelligence e data visualization.

L'obiettivo di questo lavoro, come sopra anticipato, non è certo una critica delle manchevolezze ma mira piuttosto ad agevolare il rafforzamento della *findability* degli articoli, in particolare in prospettiva DEI, e rientra in un più ampio lavoro di ricerca in corso su alcune riviste scientifiche italiane di ambito umanistico.

Sono stati analizzati tutti gli articoli presenti sulla rivista *Umanistica Digitale* dal primo numero del 2017 al numero 18 del 2024. In totale si tratta di 205 articoli, nei quali sono comprese le introduzioni e le recensioni per quanto riguarda gli autori,² e nel caso in cui fossero presenti, le parole chiave significative.³

3. RISULTATI

Un primo dato da cui partire riguarda gli autori in una prospettiva di genere. In totale sono stati individuati 327 autori. Di questi, 162 sono donne e 165 uomini. L'analisi degli autori vede perciò una sostanziale parità fra i generi se si considerano tutti i coautori dell'articolo, mentre se si considera solo il primo (o l'unico) autore di un paper, le percentuali vedono le donne con una maggior presenza (54% circa).

Come sovente accade, sia nelle riviste sia nelle banche dati di altro genere, si è riscontrata anche in questo caso una mancanza di uniformità nel modo in cui vengono indicate le istituzioni di affiliazione, per cui non è stato possibile elaborare in questo senso i dati. Si evidenzia la necessità di trovare degli identificatori anche per gli enti per favorire la ricerca e l'individuazione univoca dei medesimi con maggior facilità. Le banche dati commerciali lavorano a uniformare le affiliazioni e presentano dati più precisi a riguardo, tuttavia, il dato non risultava particolarmente rilevante per le finalità del lavoro e si è deciso che fosse sufficiente individuare le affiliazioni alle università e non ai dipartimenti. Anche in questo caso si sono uniformate le modalità citazionali, prima utilizzando OpenRefine e quindi con una ulteriore verifica di eventuali match non riscontrati dal software.

² Secondo i dati di Scopus gli articoli rappresentano comunque il 95,9% delle tipologie in *Umanistica Digitale*.

³ Sono state ovviamente escluse le parole chiave “recensione” o “review.”

Con pochissime eccezioni non sono presenti keywords riconducibili all'ambito DEI.

I dati analizzati con OpenRefine sostanzialmente coincidono con quanto si può rilevare in Scopus e SciVal, perlomeno dal punto di vista quantitativo. Ciò che differisce è l'organizzazione delle parole chiave che, nei due prodotti commerciali, sono standardizzate e fortemente strutturate. Per quanto riguarda *Umanistica Digitale*, la banca dati SciVal presenta 98 topics e 65 topic clusters, in aree disciplinari concentrate ovviamente in prevalenza tra le scienze umanistiche e le scienze sociali.⁵

SciVal utilizza il suo Fingerprint Engine per attribuire delle keyphrases tramite text mining e varie tecniche di Natural Language Processing applicate ad articoli, abstract e dati bibliografici e confrontate con alcuni thesauri di riferimento, e per l'indicizzazione in Topics, Topic Clusters e Research Areas.

Per fare un esempio, usiamo il Topic *Digital; Critical Edition; Semantic Web* che è il più rilevante in termini di preminenza globale (su base citazionale naturalmente) rispetto alla rivista *Umanistica Digitale* (include dodici articoli) e viene definito nella banca dati come relativo a "the development and impact of digital scholarly editions in humanities research. It encompasses the use of digital methods to enhance access, preservation, and analysis of historical texts and manuscripts". Il Topic presenta le seguenti keyphrases: Digital Edition; Digital; Archival Document; Scholarly edition; Online; Critical edition; Semistructured Data; Review; Scholars; Collaborative; Software; Development Tool; Online Edition; Development Version; Textual Culture; Search Engine.

In sintesi, in aggiunta alle keywords presenti sulla rivista che vengono attribuite dall'autore o dagli autori dell'articolo, e che sono sempre presenti su Scopus, SciVal presenta diversi metodi di indicizzazione propri, arricchendo significativamente il campo semantico di ciascun articolo tramite molteplici keywords o keyphrases.

L'attenzione con cui l'editore della banca dati lavora sulle keywords ne dimostra l'importanza per la possibilità di scoperta e condivisione dei risultati della ricerca scientifica pubblicati. L'uso di diversi metodi massimizza *discoverability* e *findability* dei contenuti e consente innumerevoli efficaci rielaborazioni dei dati, soprattutto alla luce delle tecniche di text and data mining e delle applicazioni crescenti dell'intelligenza artificiale.

Per tornare alla rivista di nostro interesse, l'analisi sui dati di *Umanistica Digitale* in Scopus e SciVal conferma la quasi nulla presenza di tematiche legate a diversità, equità e inclusione. Questo dato viene confermato anche dalla mappatura degli SDG (Sustainable Development Goals) applicata agli articoli della rivista sempre in SciVal. Solo ventidue articoli sono riconducibili agli Obiettivi di sviluppo sostenibile dell'Agenda 2030 (Nazioni Unite 2015). Di questi uno è ricondotto all'SDG 10 – Ridurre le disuguaglianze, quindi direttamente ed esplicitamente legato ai temi DEI, e quattro articoli sono ricondotti all'SDG 16 – Pace, giustizia e istituzioni sane, un goal non direttamente e necessariamente rilevante per DEI ma in realtà i quattro articoli lo sono, se li si va a consultare.

4. CONCLUSIONI

Gli editori sono attenti e quindi in molti casi offrono indicazioni precise agli autori e di norma utilizzano un soggetto o un thesaurus per l'indicizzazione dei contenuti bibliografici digitali. Il modo in cui i thesauri coprono le tematiche DEI sarebbe da indagare ma esula dagli obiettivi di questo contributo. Senz'altro vi sono ormai diversi studi che si soffermano sui temi DEI nella comunicazione scientifica, alcuni dei quali sono stati riportati nell'introduzione. Anche l'Agenda 2030 e l'attribuzione degli SDG ai contenuti scientifici aiuta a individuarli.

L'assenza di parole chiave relative a DEI non implica necessariamente la scarsa attenzione a queste tematiche dagli articoli nelle riviste. Nonostante la loro importanza di "connettori tra l'articolo e il lettore" spesso gli autori sottovalutano l'importanza delle parole chiave, come indicano diversi degli studi citati, e allo stesso tempo in molti casi non vi sono linee guida e indicazioni specifiche da parte degli editori e dei manager delle riviste. Questo uso poco organizzato può avere ricadute sull'attribuzione di significato e sulla ricerca di informazione in generale, e in ambito DEI in particolare. La poca attenzione attribuita dagli autori alle parole chiave non permette di evidenziare tali specificità. L'assenza di keywords legate a DEI potrebbe indicare l'assenza di tali argomenti oppure la non precisa attribuzione di parole chiave. L'uso di

⁵ I *Topics*, come spiegato nella banca dati, sono raccolte di pubblicazioni con un interesse intellettuale comune, basato su citazioni dirette, che rappresentano campi di ricerca a livello globale, costantemente in evoluzione in base ai cambiamenti nelle prospettive di ricerca. Una pubblicazione può appartenere a un solo Topic, e un Topic può appartenere a un solo Topic Cluster. La copertura temporale della rivista in Scopus (e quindi in SciVal) va dal 2019 al 2023.

keywords non corrette può essere fuorviante oppure non permettere il recupero dell'articolo; l'uso di keywords che non evidenzino i temi DEI oscura questi aspetti anche se trattati all'interno di un articolo. Tra i punti di attenzione va di certo ricompreso il modo in cui viene scritta la keyword fa la differenza. Per esempio, "person with disability" ha una valenza diversa di "disabled person" in quanto nel primo caso si accentua l'enfasi sulla persona mentre nel secondo caso sulla disabilità, così come "persona con disabilità" ha un'implicazione diversa da "disabile", soprattutto se estrapolata dal contesto (Thompson, 2020). Un tema cruciale è poi quello linguistico, al quale bisognerebbe senz'altro prestare maggiore attenzione in un contesto scientifico colonizzato dalla lingua inglese (Horvath, A., Wagner, C. Wrisley, D.J., et al. 2024), in quanto la lingua è uno strumento fortemente identitario per una comunità e la convergenza verso un'unica lingua nella comunicazione scientifica non favorisce la diversità. Tutti questi aspetti sono essenziali nella gestione automatizzata tramite text mining dei contenuti di una rivista o di una banca dati. In generale queste sono questioni alle quali gli editori stanno attribuendo considerazione crescente (APA 2021, Cuellar 2021, Dewidar 2022, Hinton & Lambert 2022; Marabelli & Chan 2024) poiché la centralità di questi temi nella società è di fatto crescente, anche alla luce dell'Agenda 2030, degli effetti della pandemia e di episodi come la morte di George Floyd o il movimento #metoo (Marabelli & Chan 2024). Tali eventi hanno accresciuto la sensibilità verso queste tematiche, e aumentato perfino gli studi che hanno dimostrato i benefici anche in termini economici di un approccio rivolto alla diversità, all'equità, all'inclusione (McKinsey & Co. 2022). L'evoluzione degli ultimi anni ha messo in luce il lavoro ancora da compiere ma allo stesso tempo ha rappresentato un punto di svolta nelle tematiche DEI, dal quale auspicabilmente non si può tornare indietro. La proposta è quindi, per tutte le riviste che ancora non lo fanno, di fornire e delle indicazioni precise, adottare linee guida, offrire suggerimenti e uno schema di tematiche DEI in cui far rientrare le keywords, oltre ad incrementare la trattazione di questi argomenti nella rivista, magari attraverso specifiche call for papers così come in questo convegno.

BIBLIOGRAFIA

- APA. (2021). Equity, Diversity, and Inclusion Toolkit for Journal Editors, <https://www.apa.org/pubs/authors/equity-diversity-inclusion-toolkit>.
- Cuellar, NG. (2021). Diversity, Equity, and Inclusivity in Peer-Reviewed Journals. *Journal of Transcultural Nursing*, 32(5), 449-450. doi:10.1177/10436596211034902
- Dewidar, O., Elmetekawy, N. & Welch, V. (2022). Improving equity, diversity, and inclusion in academia. *Research Integrity and Peer Review*, 7. <https://doi.org/10.1186/s41073-022-00123-z>
- Ferreira Garcia, D.C., Chaves Gattaz, C., & Chaves Gattaz, N. (2019). The Relevance of Title, Abstract and Keywords for Scientific Paper Writing, *Revista de Administração Contemporânea*, 3. <http://doi.org/10.1590/1982-7849rac2019190178>.
- Hartley, J., & Kostoff, Ronald D. (2003). How useful are 'key words' in scientific journals? *Journal of Information Science*, 29, 433-438.
- Hinton, A. Jr., & Lambert, M.W. (2022). Moving diversity, equity, and inclusion from opinion to evidence. *Cell Reports Medicine*, 3(4), 100619. <https://doi.org/10.1016/j.xcrm.2022.100619>
- Horvath, A., Wagner, C. Wrisley, D.J., et al. (2024). Multilinguality in Action: Towards Linguistic Diversity and Inclusion in Digital Humanities. *magazén*, 5(2). DOI: 10.30687/mag/2724-3923/2024/02/001/
- Marabelli, M., Chan, Y.E. (2024). The strategic value of DEI in the information systems discipline. *The Journal of Strategic Information Systems*, 33(1) <https://www.sciencedirect.com/science/article/pii/S0963868724000052>

- McKinsey & Company (2022). What is diversity, equity, and inclusion?
<https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-diversity-equity-and-inclusion>
- Morriello, R., & Sardo, L. (2023). Diversity management e prospettive DEI in biblioteca. *Biblioteche oggi Trends*, 9(2), 33-42. <http://doi.org/10.3302/2421-3810-202302-033-1>
- Nazioni Unite (2015). Obiettivi per lo Sviluppo Sostenibile, <https://unric.org/it/agenda-2030/>
- Raju, J. (2020). Diversity, Inclusion, and Social Justice in the Information Context: Global South Perspectives, *The International Journal of Information, Diversity, & Inclusion*, 4(3/4). DOI:10.33137/ijidi.v4i3/4.34974
- Thompson, K.M. (2020). Socializing Engagement: From Words to Action, *The International Journal of Information, Diversity, & Inclusion*, 4(1). DOI: 10.33137/ijidi.v4i1.33251