

# La ricerca in ascolto, come l'intelligenza artificiale può aiutare l'open access. Un progetto Inapp

Chiara Carlucci e Valeria Cioccolo

*INAPP – Istituto nazionale per l'analisi delle politiche pubbliche*

La forza della condivisione, in particolare per i prodotti della ricerca finanziata da fondi pubblici, e la sua disponibilità attraverso i principi dell'open access e l'open science è emersa significativamente durante la recente crisi pandemica e si concretizza attraverso ambienti digitali e archivi aperti che ne garantiscono la fruizione da parte di tutti i potenziali beneficiari: stakeholders, cittadini, società civile. L'INAPP (Istituto nazionale per l'analisi delle politiche pubbliche) svolge ricerche e analisi nell'ambito dei temi del lavoro, della formazione, delle politiche sociali e applica da molti anni i principi dell'open access mettendo a disposizione il proprio know-how tecnico-scientifico in un archivio istituzionale denominato [Inapp OA](#).

Per massimizzare la visibilità e la ricercabilità delle proprie risorse, l'INAPP ha avviato un progetto sperimentale nel quale il sistema di ricerca dell'open archive è stato implementato attraverso una rete neurale di intelligenza artificiale (AI) che potesse massimizzare i risultati di ricerche effettuate dagli utenti colmando lo scarto esistente tra i contenuti restituiti dal sistema e quelli potenzialmente disponibili. L'intelligenza artificiale così implementata lavora sia sui metadati sia sui full text disponibili nell'open archive indipendentemente dal loro formato (pdf, jpeg, tiff, png), creando una base dati "raw" formata da stringhe testo e keywords. A questa operazione si affianca il NER (Named-entity recognition) o individuazione delle 'entità nominate' per cui ogni parola significativa (persone, luoghi ed altre entità) viene progressivamente taggata e riconosciuta dall'AI che "impara" e quindi restituisce anche questa tipologia di informazioni nei risultati di ricerca. Le

simulazioni di ricerca effettuate su un utente 0 con e senza la soluzione di AI per misurarne il funzionamento, hanno mostrato come i risultati cambiassero e si arricchissero in presenza/assenza tale upgrade in AI.

Viene in tal modo sensibilmente migliorata la capacità di risposta del sistema, perché la rete neurale artificiale 'interpreta', 'traduce' e 'corregge' le parole utilizzate superando le limitazioni tipiche delle piattaforme di open archive. Dal lato back-office, inoltre, il sistema di AI coadiuva il lavoro di bibliotecari/documentalisti e professionisti della PA perché aiuta a individuare il panorama significativo dei temi di maggior interesse e delle parole del linguaggio naturale maggiormente utilizzate nelle ricerche, arricchendo così di nuovi metadati la catalogazione sia descrittiva che semantica derivanti dall'aggregazione delle informazioni usate in fase ricerca dagli utenti.

Sebbene questo progetto sia ancora in fase sperimentale, gli elementi di innovatività e di applicazione sono molteplici. Tra i più significativi, la possibilità di diffusione e integrazione con ambienti semantici, competenze e aree di applicazione diversi grazie alla replicabilità e scalabilità delle tecnologie applicate. L'ambizione è mettere la ricerca pubblica in ascolto delle esigenze reali di ricerca dei propri utenti in modo di poter consolidare il rapporto di fiducia tra istituzioni e società civile. L'iniziativa nel 2021 è stata inserita da [Open Gov](#) tra le nove finaliste della categoria del Premio Open Gov nell'ambito della sezione dedicata ai progetti delle amministrazioni pubbliche per la categoria 'Trasparenza e Open Data'.

## **Topic**

Le pubblicazioni scientifiche - comunicazione scientifica

## **Parole chiave:**

Comunicazione scientifica; Intelligenza artificiale; Open access

### #OpenAccess

La crisi pandemica ha fatto emergere significativamente che la condivisione open di informazioni di enti di ricerca e amministrazioni pubbliche è la migliore fonte cui il cittadino possa attingere

### #Inapp.OA

- Dal 2009 Inapp persegue una politica di full open access: disponibilità di pubblicazioni e dati della ricerca
- Il repository Inapp.OA è luogo di raccolta unico per l'ente e bacino informativo-documentale interrogato dal sito web istituzionale per popolare le sezioni pubblicazioni/eventi/news
- Repository e sito condividono la medesima base di indicizzazione semantica, utilizzando i descrittori del Thesaurus Inapp

### #You

- Inapp si è messo in "ascolto" dei bisogni informativi degli utenti. È emerso uno scarto tra le risorse documentali in open archive e le ricerche effettuate dagli utenti (monitoraggio degli insights del sito e derivanti dal motore di ricerca dell'open archive)
- Per colmare questo gap si è pensato di sperimentare l'applicazione di soluzioni di AI (Artificial Intelligence Intelligenza Artificiale) e machine learning a risorse disponibili in Inapp.OA

### #DeepWeb

Nonostante i progressi degli ultimi vent'anni (open access, open science) ancora troppe risorse informative sono solo virtualmente disponibili perché non facilmente rintracciabili per gli utenti

### #IntelligenzaArtificialeBERT

STEP BY STEP: COME OPERA NELL'OPEN ARCHIVE

- Intelligenza Artificiale in Inapp OA:** è stata applicata al repository INAPP.OA la rete neurale artificiale denominata **BERT**
- Base dati "raw":** sono stati analizzati e estratti contenuti, blocchi di testo e frasi poi analizzati singolarmente per escludere stringhe con contenuto minimo di informazioni
- Addestramento:** BERT ha eseguito un periodo di "addestramento" specifico sulla base dati e sul sistema documentale Inapp, analizzando i metadati, i concetti semantici del thesaurus e i full text contenuti nell'open archive utilizzando strumenti di analisi statistica, algoritmi di conversione e sistemi di machine/deep learning. Per il riconoscimento delle "Entità" significative nominate nei documenti (Persone, Luoghi geografici ecc.) alla rete neurale utilizza un dataset in cui sono taggate le parole significative e la classe di appartenenza

**RESTITUZIONE:** Il sistema così arricchito restituisce in risposta alle ricerche dell'utente una scheda con:

- Sommarizzazione** del testo: informazioni e i concetti più significativi dei documenti in relazione alla domanda posta dall'utente e lista di frasi estratte dai testi, riordinate per importanza
- Parole chiave:** lista dei termini più significativi rispetto al tema richiesto in fase di ricerca
- Entità nominate (NER - Named Entry Recognition):** lista di nomi di persone, luoghi ed altre entità significative presenti nei documenti

### #SviluppiFuturi

L'applicazione sperimentata **migliora l'accesso all'informazione prodotta dall'istituzione**, permettendo all'utente di acquisire risorse aggiuntive rispetto alla ricerca effettuata con risultati ben leggibili e strutturati. Aiuta il sistema documentale dell'ente con **dati statistici utili al monitoraggio e all'implementazione** del sistema stesso.

Come **prospettive** prossime:

- Ricerca** interlocutori negli altri enti di ricerca che potrebbero adottare, perché già in possesso di open archive, soluzioni ed approcci analoghi.
- Investire** sull'interoperabilità semantica che sfrutta i metadati per arrivare alla struttura ontologica di riferimento rispetto alle risorse informative.
- Operare** a livello di collegamento tra i dati (linked data) per costruire una rete di risorse collegate con dataset esterni favorendo una crescita esponenziale del patrimonio e delle informazioni iniziali.
- Promuovere** e/o individuare community che condividano l'interesse all'apertura del patrimonio pubblico, e la sua disponibilità

**SCHEDA DI RESTITUZIONE: RISULTATI DI RICERCA IN INAPP.OA ARRICCHITI DA BERT**

- Parole chiave
- Sommarizzazione
- Entità nominate (NER)

C. Carlucci, V. Cioccolo, *Inapp la ricerca che ascolta. Come l'intelligenza artificiale può aiutare l'open access*, Poster in "Biblioteche per l'Open Science". GenOA week 2022, 7-11 novembre 2022



## Biblioteche per l'Open Science: Sessione Poster

Ultimo aggiornamento 17 Ottobre 2022

### Biblioteche per l'Open Science: Sessione Poster



openscience@unige.it



Chi siamo  
Università di Genova

Sistema Bibliotecario di Ateneo

Area Ricerca

CeDIA

Crediti