

Programma di formazione

Titolo

Estensione del software di OpenCitations Meta per l'assegnamento di nuovi identificativi per numeri, volumi, e articoli di riviste

Responsabile scientifico

Professor Silvio Peroni <silvio.peroni@unibo.it>, Dipartimento di Filologia Classica e Italianistica, Università di Bologna / Direttore di OpenCitations, che può essere contattato per ulteriori informazioni.

Obiettivi

OpenCitations (<http://opencitations.net>) [1] è una infrastruttura Open Science che mette a disposizione una grossa mole di metadati bibliografici e dati citazionali accademici, di qualità e copertura tali da competere con servizi proprietari, come Web of Science e Scopus. OpenCitations è no-profit e tutti i suoi servizi sono completamente gratuiti. Essa è gestita dal Research Centre for Open Scholarly Metadata dell'Università di Bologna (<https://openscholarlymetadata.org>).

Negli ultimi tre anni all'interno di OpenCitations è stato sviluppato e testato un nuovo software (lanciato nel dicembre 2022) per creare una nuova raccolta, chiamata OpenCitations Meta (<https://opencitations.net/meta>). I metadati esposti da OpenCitations Meta includono i metadati bibliografici utilizzati per descrivere informazioni di base di una risorsa bibliografica. In particolare, conserva gli identificatori delle risorse bibliografiche (ad esempio, DOI, PMID, ISSN e ISBN), il titolo, il tipo, la data di pubblicazione, le pagine e il luogo della risorsa (con i numeri di volume e di fascicolo se il luogo è una rivista). Inoltre, OpenCitations Meta contiene metadati riguardanti i principali attori coinvolti nella pubblicazione di una risorsa bibliografica, ovvero gli autori, gli editori e gli editori, ogni attore può essere caratterizzato con altri identificatori (ad esempio ORCID) se disponibili.

Tutte le risorse bibliografiche incluse in OpenCitations Meta derivano da fonti/risorse esistenti (attualmente tali risorse includono Crossref, DataCite e PubMed). In particolare, esiste un record specifico per ogni entità (che cita o che viene citata) coinvolta in ogni citazione inclusa negli Indici di OpenCitations [2] - un indice di citazioni aperte contenente tutte le entità che citano e sono citate, identificate da identificatori persistenti usati dalle fonti (cioè DOI o PMID), che sono coinvolti nelle citazioni.

Uno dei principali scopi di OpenCitations Meta è quello di gestire il problema posto dalle risorse bibliografiche che non dispongono di identificatori persistenti, associando a ciascuna di esse un nuovo identificatore persistente locale - un OpenCitations Meta Identifier (OMID). In questo modo, OpenCitations Meta può contenere i metadati di qualsiasi pubblicazione accademica, senza la necessità obbligatoria che un identificatore persistente esterno sia fornito dalla fonte dei metadati (ad esempio, DOI).

In linea con questo scopo, l'obiettivo principale di questo lavoro è quello di migliorare ulteriormente i dati di OpenCitations Meta dati tramite la definizione di nuovi OMID per i numeri e i volumi delle riviste e l'associazione e integrazione di ulteriori identificatori usati da servizi esterni quali OpenAlex (<https://openalex.org>) per la definizione degli articoli delle riviste e delle riviste stesse.

Piano di attività

La Borsa di Ricerca avrà durata di 6 mesi a partire da Maggio 2023, con l'opzione di una futura estensione a termine del periodo di lavoro. Il Borsista di Ricerca lavorerà direttamente con il Professor Silvio Peroni nel contesto del Research Centre for Open Scholarly Metadata, presso il Dipartimento di Filologia Classica e Italianistica dell'Università di Bologna (Italia). Il Centro di Ricerca è un ambiente vivo e stimolante, ed è atteso che il Borsista di Ricerca fornisca contributi personali centrali alle attività di OpenCitations. Il lavoro a distanza può essere possibile se strettamente necessario, ma altrimenti la presenza di persona nel Centro di Ricerca è preferibile.

Durante il periodo di lavoro è prevista una fase iniziale introduttiva e conoscitiva dei software ed in generale della infrastruttura OpenCitations. Dopodiché il lavoro del Borsista di Ricerca può essere organizzato e riassunto in questi punti:

- 1) Registrare OMID come identificatore persistente nei servizi appropriati (ad esempio, <https://identifiers.org>);
- 2) Utilizzare OMID per fornire identificatori persistenti ai contenitori di riviste (cioè fascicoli e volumi) di tutti gli articoli di riviste coinvolti nelle citazioni incluse negli Indici di OpenCitations;
- 3) Includere il nuovo identificativo OMID per tutti i fascicoli e volumi nel (a) processo di produzione dei dump di OpenCitations Meta e (b) nelle risposte fornite dal servizio di API REST;
- 4) Estendere il software di OpenCitations Meta per gestire e tenere traccia degli identificatori di OpenAlex (<https://openalex.org/>);
- 5) Migliorare i dati di OpenCitations Meta per includere anche gli identificatori di OpenAlex per gli articoli di riviste e le riviste incluse in OpenCitations Meta;
- 6) Integrare all'interno del processo di ingestione attuale di OpenCitations Meta un plugin che permetta un allineamento agli identificatori di OpenAlex. Questo processo deve avvenire in maniera sistematica, ovvero ogni qualvolta che viene aggiunto un nuovo articolo di rivista o una nuova rivista a OpenCitations Meta;
- 7) Scrivere e pubblicare una documentazione appropriata per descrivere i risultati ottenuti da questo lavoro.

La nuova versione di OpenCitations Meta (con licenza CC0) includerà i dati aggiuntivi derivanti dal processo menzionato sopra, ovvero quello di includere (a) dei OMID per i numeri e i volumi delle riviste e (b) identificatori usati da OpenAlex per gli articoli delle riviste e le riviste stesse.

Mentre il professor Peroni dirigerà e supervisionerà il lavoro, il Borsista di Ricerca avrà la responsabilità di gestire in modo autonomo e sistematico queste attività.

Requisiti

Tutti/e i/le candidati/e devono avere eccellenti abilità come programmatori/trici e, come valore aggiunto, devono essere in grado di parlare, scrivere, e presentare verbalmente a conferenze in un buon inglese. Esperienze dimostrabili di programmazione in Python e utilizzo dei più comuni librerie Python, e sistemi di versionamento basati su Git (in particolare GitHub) sono fortemente desiderabili. In più, è altresì fortemente desiderabile che il/la candidato/a abbia una forte e dimostrabile attitudine verso la Scienza Aperta e la capacità di lavorare in gruppo. Conoscenze dimostrabili nelle tecnologie del Web Semantico, Linked Data e tecnologie Web in generale sono elementi favorevoli per la candidatura.

I requisiti minimi formali per la posizione sono il possesso di una Laurea Magistrale o equivalente. Il candidato deve avere un'esperienza adeguata e dimostrabile come programmatore, comprovata dai documenti da allegare in fase di domanda. La candidatura (in Italiano o in Inglese) deve almeno includere un Curriculum Vitae completo di informazioni riguardanti attività scientifico-professionali e relative alla produttività scientifica. Eventuali lettere di raccomandazione sono opzionali, ma fortemente consigliate.

L'Università di Bologna è un'istituzione che da pari opportunità di impiego, e la selezione per questa posizione verrà fatta esclusivamente sul merito.

Riferimenti

1. Peroni, S., & Shotton, D. (2020). OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies*, 1(1), 428–444.
https://doi.org/10.1162/qss_a_00023
2. Heibi, I., Peroni, S., & Shotton, D. (2019). Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. *Scientometrics*, 121(2), 1213–1228.
<https://doi.org/10.1007/s11192-019-03217-6>