

Programma di formazione

Titolo

Gestione, aggiornamento ed evoluzione delle Semantic Publishing and Referencing (SPAR) Ontologies

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 (<https://opencitations.net>) [1] è una infrastruttura Open Science supportata dalla comunità scientifica e accademica che fornisce informazioni sulla ricerca aperte, accessibili e riutilizzabili per favorire processi trasparenti e riproducibili a supporto della ricerca, delle decisioni politiche e della reperibilità del dominio accademico. Questo si realizza attraverso la raccolta e la pubblicazione aperta di metadati accurati e completi che descrivono le pubblicazioni accademiche e le relative citazioni che le collegano. OpenCitations rende disponibili i dati delle citazioni e i metadati bibliografici in modo che chiunque, ovunque nel mondo, possa utilizzarli per qualunque scopo. 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 sono stati sviluppati nuovi software utilizzati per creare due collezioni di metadati bibliografici e citazionali, ovvero OpenCitations Meta [2] e OpenCitations Index [3]. I metadati esposti da OpenCitations includono, oltre che ai collegamenti citazionali, 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). Sono inclusi anche i 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. Tutti i dati in OpenCitations derivano da fonti/risorse esistenti (attualmente tali risorse includono Crossref, DataCite, PubMed, OpenAIRE e JaLC).

Uno dei principali scopi di OpenCitations è quello di mantenere una collezione di ontologie, the Semantic Publishing and Referencing (SPAR) Ontologies (<http://www.sparontologies.net>) [4], che sono state usate fin dal 2010 dall'infrastruttura per descrivere i dati che essa colleziona e renderli *machine-actionable* dalle macchine, oltre che essere usate sistematicamente anche dalla OpenCitations Ontology e da altri progetti a carattere internazionale tra cui il Scientific Knowledge Graph – Interoperability Framework (SKG-IF, <https://skg-if.github.io>). Recentemente, l'intero sistema legato alle SPAR ontologies è stato migrato formalmente sulla nuova infrastruttura tecnologica di OpenCitations. Per questa ragione, esiste la necessità di

aggiornare tutto l'impianto formale delle ontologie, sistemando alcuni problemi concettuali identificati negli anni scorsi, e di fornire l'appropriato contesto documentale atto a rendere le ontologie facilmente fruibili sia da umani sia da macchine.

Piano di attività

La Borsa di Ricerca avrà durata di 8 mesi a partire da Marzo 2026. 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, in particolare relativamente alle SPAR Ontologies. 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 del contesto applicativo. Dopodichè il lavoro del Borsista di Ricerca può essere organizzato e riassunto in questi punti:

- 1) Sviluppare nuove e/o arricchire esistenti ontologie SPAR per la descrizione dei diversi ambiti dell'ecosistema accademico.
- 2) Raccogliere ed organizzare un corpus adeguato di evidenze che dimostrino in quanti differenti contesti siano state usate le SPAR Ontologies, e definire un facile sistema per aggiornare in futuro questi riferimenti
- 3) Ridisegnare e ridefinire, mediante lo sviluppo e il riuso di sistemi computazionali, la documentazione HTML di tutte le ontologie, includendo anche documentazioni puntuale per tutte le entità che esse descrivono e non soltanto dell'ontologia principale che le definisce.
- 4) Sviluppare dei sistemi che facilitino la validazione di dati conformi a queste ontologie mediante dei documento SHACL.
- 5) Partecipare attivamente alla comunità di SKG-IF per dare supporto legato alla semantica e validazione del data model.
- 6) Scrivere e pubblicare una documentazione appropriata per descrivere i risultati ottenuti da questo lavoro.

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 *ontology engineer* e programmatore/trici e, come valore aggiunto, devono essere in grado di parlare, scrivere, e presentare verbalmente a conferenze in un buon inglese. Esperienze dimostrabili di conoscenza dei linguaggi di sviluppo di artefatti semantici (RDF, RDFS, OWL, SKOS), di programmazione in Python e utilizzo dei più comuni librerie Python, e di 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 in LM43. Il candidato deve avere un'esperienza adeguata e dimostrabile come ontology engineer e 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. Massari, A., Mariani, F., Heibi, I., Peroni, S., & Shotton, D. (2024). OpenCitations Meta. *Quantitative Science Studies*, 5(1), 50–75. https://doi.org/10.1162/qss_a_00292
3. Heibi, I., Moretti, A., Peroni, S., & Soricetti, M. (2024). The OpenCitations Index: Description of a database providing open citation data. *Scientometrics*, 129(12), 7923–7942. <https://doi.org/10.1007/s11192-024-05160-7>
4. Peroni, S., & Shotton, D. (2018). The SPAR Ontologies. In L. Rutkowski, R. Scherer, M. Korytkowski, W. Pedrycz, R. Tadeusiewicz, & J. M. Zurada (Eds), *The Semantic Web – ISWC 2018* (Vol. 10842, pp. 119–136). Springer. https://doi.org/10.1007/978-3-030-00668-6_8