

Programma di formazione

Titolo

Sviluppo dell'infrastruttura computazionale di OpenCitations

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 stata creata come una **infrastruttura aperta e completamente gratuita** 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**. OpenCitations è gestita dal Research Centre for Open Scholarly Metadata dell'Università di Bologna (<https://openscholarlymetadata.org>). Inoltre, è impegnata nella promozione delle citazioni aperte, in particolare nel suo ruolo di membro fondatore della Initiative for Open Citations (I4OC, <https://i4oc.org>) della Initiative for Open Abstracts (I4OA, <https://i4oa.org>).

Alla fine del 2019, OpenCitations è stata selezionata dalla Global Sustainability Coalition for Open Science Services (SCOSS, <https://scoss.org>) per il loro secondo ciclo di supporto per la raccolta fondi, visto che OpenCitations è ben allineata con gli obiettivi propri della Scienza Aperta ed offre un servizio innovativo. Inoltre, OpenCitations prende parte, attraverso l'Università di Bologna, ad una serie di iniziative internazionali tra cui OpenAIRE-Nexus, un progetto Horizon 2020 finanziato dalla Commissione Europea (<https://cordis.europa.eu/project/id/101017452>) che ha l'obiettivo di portare, all'interno della **European Open Science Cloud (EOSC)**, quattordici nuovi servizi (tra cui OpenCitations) focalizzati sullo sviluppo e promozione della Scienza Aperta.

La mole e la complessità di dati messi a disposizione da OpenCitations aprono numerosi problemi di scalabilità e ottimizzazione. È necessario quindi studiare e personalizzare le più recenti

tecniche di gestione di Big Data per creare una base di conoscenza su cui costruire servizi scalabili, duraturi e facilmente personalizzabili. L'obiettivo della Borsa di Ricerca, infatti, è quello di studiare le limitazioni e i miglioramenti da apportare all'attuale infrastruttura computazionale in modo da venire incontro alle esigenze di estensione di servizi e di dati richieste dai vari stakeholders – che prevedono l'erogazione di servizi Web, la gestione di basi di dati distribuite ed eterogenee, e processi di conversione e ingestione dati.

Il lavoro da svolgersi per questa posizione è di natura non commerciale.

Piano di attività

La Borsa di Ricerca avrà durata di 4 mesi a partire da Giugno 2021, eventualmente rinnovabile per ulteriori 8 mesi. 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 per la situazione relativa alla pandemia di COVID-19, ma altrimenti la presenza di persona nel Centro di Ricerca è preferibile.

Durante il primo mese, il Borsista di Ricerca verrà introdotto/a a tutti i sistemi hardware e software usati da OpenCitations per fornire i propri servizi, così da capire lo stato corrente dell'infrastruttura tecnologica. Nei restanti mesi, il Borsista di Ricerca assumerà la responsabilità di gestione e sviluppo di questi sistemi, proponendo soluzioni innovative basate su pratiche di ricerca aggiornate, ove necessario, in modo da permettere l'integrazione e l'estensione dei servizi offerti da OpenCitations. Questi sistemi includono:

- Un server fisico principale, che fornisce tutti i servizi di OpenCitations;
- Le basi di dati di tutti i dataset di OpenCitations, che includono l'OpenCitations Corpus [2], gli OpenCitations Indexes [3], il Open Biomedical Citations in Context Corpus [4], e il futuro OpenCitations Meta.

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 nell'usare e gestire server Linux (in particolare Debian e Raspbian), server Web che implementano FastGCI (in particolare lighttpd), basi di dati basate su grafi (in particolare Blazegraph), programmazione in Python e utilizzo dei più comuni Python Web Application Framework (in particolare web.py e Flask), 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 Triennale o 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. Peroni, S., Shotton, D., & Vitali, F. (2017). One Year of the OpenCitations Corpus. In C. d'Amato, M. Fernandez, V. Tamma, F. Lecue, P. Cudré-Mauroux, J. Sequeda, C. Lange, & J. Heflin (Eds.), *The Semantic Web – ISWC 2017 (Lecture Notes in Computer Science Vol. 10588, pp. 184–192)*. Springer. https://doi.org/10.1007/978-3-319-68204-4_19, disponibile in open access all'URL <https://w3id.org/people/essepuntato/papers/oc-iswc2017.html>
3. 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>

4. Daquino, M., Peroni, S., Shotton, D., Colavizza, G., Ghavimi, B., Lauscher, A., Mayr, P., Romanello, M., & Zumstein, P. (2020). The OpenCitations Data Model. In J. Z. Pan, V. Tamma, C. d'Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne, & L. Kagal (Eds.), *The Semantic Web – ISWC 2020* (Lecture Notes in Computer Science Vol. 12507, pp. 447–463). https://doi.org/10.1007/978-3-030-62466-8_28, disponibile in OA all'URL <https://arxiv.org/abs/2005.11981>

Research programme

Title

Developing and managing the OpenCitations computational infrastructure

Supervisor

Professor Silvio Peroni <silvio.peroni@unibo.it>, Dipartimento di Filologia Classica e Italianistica, Università di Bologna / Director of OpenCitations, from whom further information may be obtained.

Goals

OpenCitations (<http://opencitations.net/>) [1] has been established as a **fully free and open infrastructure** to provide access to global **scholarly bibliographic and citation data**, of quality and coverage to rival those from proprietary services, such as Web of Science and Scopus. OpenCitations is **not-for-profit** and all **its services are free**. OpenCitations is managed by the Research Centre for Open Scholarly Metadata of the University of Bologna (<https://openscholarlymetadata.org>). It is also engaged in advocacy for open citations, particularly in its role as a key founding member of the Initiative for Open Citations (I4OC, <https://i4oc.org>) and of the Initiative for Open Abstracts (I4OA, <https://i4oa.org>).

At the end of 2019, OpenCitations was selected by the Global Sustainability Coalition for Open Science Services (SCOSS, <https://scoss.org>) for its second round of crowdfunding support, since OpenCitations aligns well with Open Science goals and is an innovative service. In addition, OpenCitations is involved in, via the University of Bologna, a series of international initiatives including OpenAIRE-Nexus, an H2020 project funded by the European Commission (<https://cordis.europa.eu/project/id/101017452>) which aims at bringing, in the [European Open Science Cloud \(EOSC\)](#), fourteen new services (including OpenCitations) focused on the development and promotion of Open Science.

The amount and complexity of the data made available by OpenCitations opens up to several issues related to the scalability and optimization of its infrastructure. Therefore, it is crucial to study and customize recent approaches for the management of Big Data to create a knowledge

base to build scalable, long-lasting and easily-customizable services. Indeed, the goal of the Short-Term Research Fellowship is to study the current limitations and possible improvements to introduce in the OpenCitations computational infrastructure to extend services and data as required by several stakeholders – which include the provision of Web services, the management of distributed and heterogeneous databases, and data conversion and ingestion processes.

The work related to this position is non-commercial in nature.

Activity plan

The Short-Term Research Fellowship will have a duration of 4 months commencing in June 2021, that can be renewed for additional 8 months. The Short-Term Research Fellow will work directly with Professor Silvio Peroni in the context of the Research Centre for Open Scholarly Metadata at the Department of Classical Philology and Italian Studies, University of Bologna, Italy. This is a multicultural, lively and stimulating environment, and the Research Fellow will be expected to provide a key personal contribution to the OpenCitations activities. Distance working will be permitted if necessary because of the COVID-19 pandemic, but otherwise attendance in person in the Research Centre will be preferable.

During the first month, the Short-Term Research Fellow will be introduced to all the hardware and software systems currently used by OpenCitations to provide its services, so as to become familiar with the current status of the technical infrastructure. Subsequently, the Short-Term Research Fellow will assume responsibility for managing and developing these systems, with the introduction of innovative solutions coming from current research practices when necessary, so as to improve and extend OpenCitations services. They comprise:

- The main physical server providing the OpenCitations services;
- The databases storing the data of all the OpenCitations datasets, including the OpenCitations Corpus [2], the OpenCitations Indexes [3], the Open Biomedical Citations in Context Corpus [4], and the future OpenCitations Meta dataset.

While Professor Peroni will direct and supervise his/her work, the Short-Term Research Fellow will be given, and will be expected to exercise, responsibility in undertaking these tasks.

Requirements

Applicants are expected to have **excellent computer programming skills** and, as an added value, the ability to speak, write and make verbal conference presentations in good English. Demonstrable expertise in using and managing Linux servers (especially Debian and Raspbian), Web servers implementing FastGCI (especially lighttpd), graph databases (especially Blazegraph), Python programming and Python Web Application Frameworks (especially web.py and Flask), Git-based version-control systems (especially GitHub), plus a strong and demonstrable commitment to open science and team-working abilities are highly desirable. Demonstrable knowledge in Semantic Web technologies, Linked Data and Web technologies would also be very beneficial.

The minimal formal requirement for this position is a Bachelor or Master degree, or equivalent. It is expected that the successful applicant will have had substantial and demonstrable experience in computer programming, described in the documents to submit for the application. The Short-Term Research Fellowship application (written either in Italian or in English) should minimally comprise a full Curriculum Vitae containing information about prior scientific and professional jobs and about the scientific production of the candidate. Letters of recommendation are optional but highly desirable.

The University of Bologna is an equal-opportunities employer, and selection for this post will be made solely on merit.

References

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. Peroni, S., Shotton, D., & Vitali, F. (2017). One Year of the OpenCitations Corpus. In C. d'Amato, M. Fernandez, V. Tamma, F. Lecue, P. Cudré-Mauroux, J. Sequeda, C. Lange, & J. Heflin (Eds.), *The Semantic Web – ISWC 2017 (Lecture Notes in Computer Science Vol. 10588, pp. 184–192)*. Springer. https://doi.org/10.1007/978-3-319-68204-4_19, disponibile in open access all'URL <https://w3id.org/people/essepuntato/papers/oc-iswc2017.html>

3. 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>
4. Daquino, M., Peroni, S., Shotton, D., Colavizza, G., Ghavimi, B., Lauscher, A., Mayr, P., Romanello, M., & Zumstein, P. (2020). The OpenCitations Data Model. In J. Z. Pan, V. Tamma, C. d'Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne, & L. Kagal (Eds.), *The Semantic Web – ISWC 2020 (Lecture Notes in Computer Science Vol. 12507*, pp. 447–463). https://doi.org/10.1007/978-3-030-62466-8_28, disponibile in OA all'URL <https://arxiv.org/abs/2005.11981>