2018-00234 - [NEO] PhD Scholarship at NEO (EP) / Networks and Telecommunications

**Contract type:** Public service fixed-term contract  
**Level of qualifications required:** Graduate degree or equivalent  
**Function:** PhD Position  
**Level of experience:** Recently graduated

**About Inria**

Inria, the French National Institute for computer science and applied mathematics, promotes “scientific excellence for technology transfer and society”. Graduates from the world's top universities, Inria's 2,700 employees rise to the challenges of digital sciences. With its open, agile model, Inria is able to explore original approaches with its partners in industry and academia and provide an efficient response to the multidisciplinary and application challenges of the digital transformation. Inria is the source of many innovations that add value and create jobs.

**About the research centre or Inria department**

The Inria Sophia Antipolis - Méditerranée center counts 37 research teams and 9 support departments. The center's staff (about 600 people including 400 Inria employees) is composed of scientists of different nationalities (250 foreigners of 50 nationalities), engineers, technicians and administrators. 1/3 of the staff are civil servants, the others are contractual. The majority of the research teams at the center are located in Sophia Antipolis and Nice in the Alpes-Maritimes. Six teams are based in Montpellier and a team is hosted by the computer science department of the University of Bologna in Italy. The Center is a member of the University and Institution Community (ComUE) “Université Côte d'Azur (UCA)”.

**Context**

PhD Opening at Inria Sophia Antipolis, France

[https://www.inria.fr/](https://www.inria.fr/)  
[https://team.inria.fr/neo/presentation/](https://team.inria.fr/neo/presentation/)  
under the supervision of Prof. K. Avrachenkov  
e-mail: K.Avrachenkov@inria.fr  
[http://www-sop.inria.fr/members/Konstantin.Avratchenkov/me.html](http://www-sop.inria.fr/members/Konstantin.Avratchenkov/me.html)

**Assignment**

**Topic:** Adaptive crawling with machine learning techniques

**Resume:** We shall consider the problem of web crawling with limited bandwidth and computational resources. Some web sites could be crawled too frequently resulting in resource underutilisation and the other web sites could be crawled not sufficiently frequently resulting in resource overutilisation. We shall try to design an adaptive crawling algorithm based on machine learning techniques such as clustering and reinforcement learning to try to find dynamically optimal crawling frequencies based on web site classification, behaviour and changes.

The project is in the framework of the joint Inria - Qwant Research Lab.

**General Information**

- **Theme/Domain:** Networks and Telecommunications  
  Statistics (Big data) (BAP E)
- **Town/city:** Sophia Antipolis
- **Inria Center:** CRI Sophia Antipolis - Méditerranée
- **Starting date:** 4/2/18
- **Duration of contract:** 3 years
- **Deadline to apply:** 3/31/18

**Contacts**

- **Inria Team:** NEO  
  **Recruiter:** Avrachenkov Konstantin  
  konstantin.avratchenkov@inria.fr

**Conditions for application**

Application file: Applications must be submitted online on the Inria website. Collecting applications by other channels is not guaranteed.

**Defence Security:**

This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

**Recruitment Policy:**

As part of its diversity policy, all Inria positions are accessible to people with disabilities.

**Warning:** you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.
Related references:

Lefortier, D., Ostroumova, L., Samosvat, E. and Serdyukov, P.,
"Timely crawling of high-quality ephemeral new content".
In Proceedings of the 22nd ACM international conference on Information & Knowledge Management

Faheem, M., and Senellart, P.,
"Adaptive Web Crawling Through Structure-Based Link Classification".

Avrachenkov, K., and Borkar, V.,
"Whittle Index Policy for Crawling Ephemeral Content".
to appear in IEEE Trans on Control of Network Systems,

Main activities

Preparation of PhD Thesis by writing journal and conference articles, developing algorithms
with possible patent application.

Skills

Required skills: solid knowledge of mathematics;
advanced knowledge of probability and statistics is a plus;
knowledge of python is another plus.

Benefits package

- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Sports facilities

Remuneration

Duration: 36 months
Location: Sophia Antipolis, France
Gross Salary per month: 1982€ brut per month (year 1 & 2) and 2085€ brut/month (year 3)