| **2020-IPR-I-000-015570** | **Information extraction from text** |

| Position for:

1 Trainee TYPE II | Background: The Text and Data Mining Unit (I3) of the European Commission's Joint Research Centre (JRC) in Ispra, Italy, is looking for a trainee to support the JRC's Europe Media Monitoring (EMM) team.

The Europe Media Monitoring is an in-house innovation service supporting EU Institution's Policy Makers, Investigators and Analysts in their knowledge-intensive tasks by providing advanced analytical tools in the field of text mining and in particular in the field of information extraction from textual news and social media.

The trainee will contribute in the Named Entity Recognition (NER) module, extract from textual news and social media known entities and advancing geolocation capabilities.

We are looking for an enthusiastic and skilful trainee for improving our Geographical Entities Recognition. He/she will work on an import of Geonames data (https://www.geonames.org/) into our NER module and schedule a periodic update. Another interesting development will be to develop a systematic evaluation method of the results of the module.

The trainee will be experimenting novel use cases involving multilingual text mining, natural language processing, knowledge discovery, and data visualisation. He/she will contribute to develop innovative software applied to a wide range of areas within European Union's Competences, and be involved in the preparation of scientific reports. In particular, the trainee will contribute to projects involving information extraction from textual news and social media in the aim of advancing geolocation capabilities.

Required skills are (a) a Master degree (or an almost completed degree) in computer science, computational linguistic, engineering, or related areas; (b) good knowledge of SQL (Postgres/PostGIS) (c) good computer literacy and at least basic programming skills (e.g. in java, python, …); (d) good working knowledge of English; (e) ability to work independently and as part of a team.

Useful skills and experiences any of following would be an asset: (f) good knowledge in the Java programming language and in testing frameworks; (h) professional experience in artificial |

intelligence, machine learning, natural language processing, semantic text analysis, etc.

In your application, please provide clear information on your skill set, by elaborating on the above-mentioned list of requirements and by listing your level of languages and your computer / programming skills.

The JRC team: The Joint Research Centre (JRC; http://ec.europa.eu/dgs/jrc/) is the scientific-technical arm of the European Commission. The approximately 2200 JRC employees working in Ispra are from all EU countries and there are also some non-EU visitors. The working environment is multilingual, multi-cultural and multi-disciplinary.

For rules and general eligibility requirements, see the link below.

Institute Unit	Directorate *Competences* I03 – *Text and Data Mining Unit*
Duration	5 months
JRC Site Country	Ispra Italy
Rules, eligibility requirements	Trainees: https://ec.europa.eu/jrc/en/working-with-us/jobs/temporary-positions/jrc-trainees