EUROPEAN COMMISSION
DIRECTORATE-GENERAL HUMAN RESOURCES AND SECURITY
Directorate HR.AMC –Account management Centre
HR.DDG. AMC 8

2018-IPR-I-000-010098	**Multilingual Entity-centric Event Extraction**

Position for: Trainee	As the science and knowledge service of the Commission, the mission of Joint Research Centre is to support EU policies with independent evidence throughout the whole policy cycle. The JRC is located in 5 Member States (Belgium, Germany, Italy, the Netherlands and Spain). Further information is available at: http://www.jrc.ec.europa.eu **Short description of activity:** The JRC's *Europe Media Monitor* (EMM) team carries out research and development in the field of highly multilingual text mining (Language Technology; Computational Linguistics) for the purposes of media monitoring. EMM gathers an average of 300,000 online news articles per day in over 70 languages and analyses them to help its large international user community understand and use this enormous amount of media information. The *Europe Media Monitor* EMM is publicly accessible and widely used. The EMM team has produced over 200 international peer-reviewed publications. The team has also produced and distributes a number of highly multilingual Language Technology resources. 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 Monitor* (EMM) team in its effort to develop a **general-purpose application that is able to scan large text collections of various types in order to compute time-ordered series of open-domain events involving a target entity** such as persons or organisations. More precisely, the task focuses on: (a) identification of all occurrences of a target entity in text collections (e.g., online news, search engine results, social media), including named mentions and mentions of entities that embrace the target entity, (b) identification of event triggers (relevant verb and noun phrases) involving the target entity, (c) classification and labelling at various levels of

	abstraction of the detected events, (d) assignment of time references to the events the target entity participated in, and (e) provision of intelligent filtering tools and visualisation of the event time series. As of now, a prototype of such entity-centric event extraction tool for processing text collections in English has been built, while the future work will embrace extensions to: cover more languages, improve the overall accuracy, cover new sources of information, merge information across documents and languages, etc. In particular, Open Information Extraction and Knowledge Harvesting techniques are used to tackle multi-linguality and scalability, these ones being the two most important design criteria in this context. The EMM team develops various applications for gathering, aggregating and analysing information from a wide range of sources, including for instance online news (NewsBrief, MediSys), search engine results (OSINT Suite) and social media. Methods used are mostly hybrid: machine learning tools are used to gather evidence, learn vocabulary and patterns, but the results are usually controlled and optimised through human intervention. EMM applications are used by European Institutions, by national authorities in EU Member States, by international organisations and by the public. EMM is part of the JRC's Competence Centre on Text Mining and Analysis. The successful trainee will contribute to the further development of the entity-centric event extraction tool which will encompass adapting the tool to process new languages (acquisition of language-specific resources) and/or improving the existing ones and devising new methods for open information extraction. The trainee is also expected to contribute to writing a scientific publication on the work carried out. **Qualifications:** **Essential**: • University degree in computational/formal linguistics, computer science or related areas; • Java programming skills; • knowledge of machine learning; • good working knowledge of English (B2 level).

	Advantage: • knowledge of further foreign languages; • the proven advanced programming skills, especially in Java; • good knowledge of Language Technology-related tools and methods, in particular in the area of Information Extraction; • the proven ability to work independently and as part of a team. 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. **For general eligibility requirements, please read the rules governing the traineeship scheme of the JRC:** https://ec.europa.eu/jrc/en/working-with-us/jobs/temporary-positions/jrc-trainees
Unit /Directorate	Directorate *Competences* I03 – *Text and Data Mining Unit*
Indicative duration **Preferred starting date**	5 months As soon as possible
JRC Site **Country**	Ispra Italy
JRC contact details	**For any technical problems with your application, please contact:** HR-AMC8-RECRUITMENT-TOOLS-SUPPORT@ec.europa.eu