



2016-IPR-G-000-6713

Multilingual Text Analysis

Position for:	<u>Short description of activity:</u>
Trainee	<p>The European Commission's <i>Joint Research Centre</i> (JRC) in Ispra, Italy, is looking for a trainee to support the JRC's <i>Europe Media Monitor</i> (EMM) team with a variety of Language Technology-related tasks. EMM gathers and analyses reports from traditional and social media in dozens of languages by clustering related news items; categorising them; extracting information such as entities (persons, organisations, locations), events (who did what to whom, where and when), quotations by and about people; identifying sentiment; as well as linking related news clusters over time and across languages. Methods used are mostly hybrid: machine learning tools are used to gather evidence, learn vocabulary and rules, but the results are usually controlled and optimised through human intervention. EMM is used by European Institutions, by national authorities in EU Member States, by international organisations and by the public. The public EMM applications can be accessed at the URL http://emm.newsbrief.eu/overview.html. The EMM team also currently contributes to a United Nations effort of updating the UNISDR Terminology on Disaster Risk Reduction (DRR) by analysing term usage in English, French and Spanish DRR-related document collections. For more information, see http://www.preventionweb.net/files/45462_ba ckgroundpaperonterminologyaugust20.pdf.</p> <p>The successful trainee will carry out any of the following tasks: (a) use third-party software to carry out a terminology use study, which includes comparing occurrences of terms and their variants in English, French and Spanish; (b) gather existing definitions of important terms from the internet; (c) improve the JRC's existing entity-oriented sentiment analysis tools, then analyse large quantities of</p>

	<p>sentiment data and its change over time with the purpose of identifying opinion change patterns and trends; (d) contribute to the semi-automatic classification of entities in JRC's multilingual entity database; (e) contribute to improving the recognition of multilingual organisation names; (f) annotation of linguistic data and/or evaluation of automatic text analysis results; (g) contribute to writing a scientific publication.</p> <p><u>Qualifications:</u> <u>Essential:</u> A degree (or an almost completed degree) in computational linguistics, computer science or related areas; Programming skills; Good command of oral and written English (level B2).</p> <p><u>Advantage:</u> Knowledge of further foreign languages; Proven advanced programming skills, especially in Java; Good knowledge of Language Technology-related tools and methods; The proven ability to work independently and as part of a team.</p> <p><u>For general eligibility requirements, please read the rules governing the traineeship scheme of the JRC:</u></p> <p>https://ec.europa.eu/jrc/en/working-with-us/jobs/temporary-positions/jrc-trainees</p>
Institute/Directorate Unit	IPSC G02 Further information: http://ipsc.jrc.ec.europa.eu
Indicative duration	5 months
Preferred starting date	1 June 2016
JRC Site Country	Ispra Italy
<u>JRC contact details</u>	For any technical problems with your application, please contact: JRC-ESRA@ec.europa.eu