



2018-IPR-I-000-9856

**Entity and Sentential-level Sentiment Analysis
in the News**

<p>Position for:</p> <p>Trainee</p>	<p>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.</p> <p>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</p> <p>The <i>Joint Research Centre</i> (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.</p> <p>The working environment is multilingual, multi-cultural and multi-disciplinary. The JRC's <i>Europe Media Monitor</i> (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 <i>Europe Media Monitor</i> 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.</p> <p><u>Short description of activity:</u></p> <p>The <i>Text and Data Mining Unit</i> (I3) of 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 in its effort to improve its multilingual sentiment analysis tools, especially at sentence and entity level. 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</p>
--	---

(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 [NewsBrief](#), [NewsExplorer](#) and [MedISys](#) can be accessed freely by the general public. EMM is part of the [JRC's Competence Centre on Text Mining and Analysis](#).

As of now, the EMM team has implemented several approaches to multilingual sentiment analysis, for different text types (newspaper articles, microblogs, social media posts) and application scenarios (document level, short texts, entity-centric).

The successful trainee will help to combine the current approaches and resources and extend them when necessary to perform multilingual entity and sentence-level sentiment analysis and evaluate the system thus obtained. The trainee is also expected to contribute to writing a scientific publication on the work carried out.

Qualifications:

Essential:

- University degree (or an almost completed degree) in computational linguistics, computer science or related areas (the degree thesis has to be registered and the subject has to match with the project);
- Java programming skills;
- Good working knowledge of English. (B2 level)

Advantage:

- Experience in methods and resources for sentiment analysis and emotion detection;
- Knowledge of further foreign languages;
- Good knowledge of Language Technology-related tools and methods;
- Proven ability to work independently and as part of a team.

	<p><i>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.</i></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>
Unit /Directorate	Directorate Competences I03 – Text and Data Mining Unit
Indicative duration	5 months
Preferred starting date	As soon as possible
JRC Site	Ispra
Country	Italy
<u>JRC contact details</u>	<p>For any technical problems with your application, please contact:</p> <p>HR-AMC8-RECRUITMENT-TOOLS-SUPPORT@ec.europa.eu</p>