EUROPEAN COMMISSION
DIRECTORATE-GENERAL
JOINT RESEARCH CENTRE
Directorate B - Resources
Recruitment and Training

2014-IPR-G-000-4154	**Terminology discovery over time in the field of disaster risk management**

Position for:	**Short description of activity:**
2 Trainees TYPE II	The *Europe Media Monitor* (EMM) group at the European Commission's *Joint Research Centre* (JRC) in Ispra, Italy, is looking for two trainees to work on a project to automatically explore the development of terminology in the field of 'Disaster Risk Management' (DRM). The purpose is to give the international stakeholders in that field (e.g. the *United Nations Office for Disaster Risk Reduction* UN-ISDR) concrete and countable evidence of new concepts (terms) emerging in their field, of changing concepts and of shifts in interest over time. The study will include both scientific publications and texts produced by national and international governmental organisations working in that field. This first exploratory study will exclusively concern English language text in the field of *Disaster Risk Management*, but other languages and subject areas will be considered in case the outcome of this exploratory study is deemed concrete and useful. This work may lead to a scientific publication co-authored by the project contributors. A scenario to reach this goal of terminology discovery might consist of the following steps: (1) Manual or semi-automatic selection and collection of freely available documents covering the sub-areas of the life cycle of Disaster Risk Management (Prevention and mitigation; Preparedness; Response; Recovery and reconstruction); (2) Conversion of the various file formats (e.g. HTML, PDF, MS-Word) into a structured text format (e.g. XML); (3) Selection of suitable off-the-shelf software for the automatic extraction of terms (e.g. noun phrases);

	(4) Usage of this software and, if needed, tuning of this software to extract lists of potential terms; (5) Application of statistical methods to select the domain-specific terms and to weigh or rank them; (6) Application of statistical methods that allow to observe trends such as the detection of terms that are more frequently or more rarely used compared to previous observation periods; (7) Presentation of the results (term lists, trends) in an easy-to-understand manner; this may also include a keyword-in-context presentation of the terms, or similar. The foreseen traineeship duration is five months, starting potentially in March 2015. The working language is English. **Required qualifications:** • Mature student or post-graduate in any of the following fields (or similar): computational linguistics, computer science, library sciences, machine learning; • Knowledge of – and experience with – freely available Language Technology tools (e.g. for terminology extraction, term weighting, categorisation); • Experience with document format conversion (PDF, HTML, MS-Word etc. to text); • Sufficient programming experience to autonomously implement all necessary steps (Java preferred); • Knowledge of statistical methods for term weighing (e.g. chi-square, TF.IDF) and for automatic categorisation; • Linguistic sensitivity and an interest for terminology extraction (what is a term?; Relationships between terms); • Ability to present the project outcome in a format suitable for DRM specialists who may not be so knowledgeable of Information Technology (presentation; reporting; visualisation). • Ability to work autonomously; • Team worker; • Good working knowledge of English (level B2) plus the ability to communicate in at least one other official EU language. In your application, please state your interests

	and please provide clear information on your skill set, by elaborating on the above-mentioned list. Should you apply as a ready-made team, please nevertheless clearly state your personal skills and strengths. **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 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. The JRC's *Europe Media Monitor* (EMM) team (http://ipsc.jrc.ec.europa.eu/?id=179) carries out research and development in the field of text mining (Language Technology; Computational Linguistics) for the purposes of media monitoring. EMM gathers an average of almost 200,000 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. EMM is publicly accessible via http://emm.newsbrief.eu/overview.html.
Institute/Directorate **Unit**	IPSC G02 Further information: http://ipsc.jrc.ec.europa.eu
Indicative duration **Preferred starting date**	5 months Approx. March 2015
JRC Site **Country**	Ispra Italy
JRC contact details	**For any technical problems with your application, please contact:** JRC-ESRA@ec.europa.eu