JD ID:
Position:
Taleo Requisition ID:
CCOG:

Machine Learning Specialist
(Junior Professional Officer)

Advanced Technology Applications Center (ATAC)
Global Database Division
Global Infrastructure Sector

Category and Grade: P2

1.   Organizational Context

The position is located in the Advanced Technology Applications Center (ATAC) of the Global Database Division, Global Infrastructure Sector. The Center is responsible for exploring new technologies (artificial intelligence (AI), mainly machine learning) in the domain of Intellectual Property. AI-based applications have been developed (using open source) and are currently used in production in three areas (text/speech/image).

Machine learning techniques are used for machine translation (WIPO Translate) applied in different domains (patent, trademarks, language division etc..). Image similarity is used in the Global Brand database (soon also in industrial designs) to search for similar images. WIPO Speech-to-text is used to automatically generate transcripts of conferences.

WIPO is looking for a Junior Professional Officer who could work on improving our tools (WIPO Translate, WIPO speech-to-text, image similarity, classification etc.). The skilled computer scientist will assist our work in data-driven Machine Learning at WIPO.

Neural networks being an emerging field, we are looking for a highly motivated person skilled in Machine learning. We focus on applied research: applying the latest research development into production.

The officer will be responsible for improving the quality of our machine learning tools by cleaning/filtering the data, improving our machine learning parameters and/or techniques, and providing results. The post is located at WIPO's headquarters in Geneva.

The JPO works under the supervision of the Head of the Advanced Technology Applications Center (ATAC).

2. Duties and responsibilities

The JPO performs the following principal duties:

Projects

The WIPO Global Databases Division is working in three main areas where machine learning brings benefits for the users: speech, text and image. WIPO is especially looking for a Junior Professional Officer who could contribute to keep WIPO's "competitive advantage" in these areas. Note that the incumbent might work on one or many of these areas (depending on the skills and business requirements).

Generic duties:
- Gather and clean training data (collect data from various sources, process it in a format that can successfully be used by machine learning algorithms).
- Advise the team on ways to enrich data resources for machine learning. Explore the usage of external resources that can be used to improve the machine learning models.
- Define metrics and isolate test and development data to carefully evaluate the quality of the trained models.
- Develop proof of concept prototypes for identified technologies, demonstrating the applicability and pertinence of new Machine Learning algorithms.
- Evaluate, plan and assist in overseeing the required IT developments (possibly outsourced), testing, deployments for successful proof of concepts and transfer knowledge to implement the targeted technology into IT components usable and maintainable by the Division.
- Perform other related tasks as required.

Specific duties:
- The Junior Professional Officer will work in one of the following area (to be agreed with the team taking into account the JPO's skills):

Text (neural machine translation-NMT/classification)

- Experiment with various techniques for setting the best parameters for specific language pairs, specific domains, and for combining various input source corpora (mixing in-domain small corpus with larger out-of-domain corpus), mixing multi languages in a single NMT model etc.
- Work on specific tools allowing a better integration of NMT in the user environment (batch translating texts/ documents/ HTML pages).
- Pre- and Post-process texts in different languages to improve translation quality (especially for Japanese).
- Collect and clean parallel sentences (e.g.: filter/clean patent titles and abstracts, align full texts using patent priority data, align documents at the sentence level, etc.).

- Work on automatic post-editing tasks: learn recurrent errors from human post editions (or from user feedback) to correct NMT output.
- Explore the feasibility of reusing and improving Large Language Models for various applications, including general text related tasks such as summarization, editing and drafting.
- Investigate and improve the application of generative AI based solutions at WIPO, including technical survey, feasibility and development of prototypes.

Images (search/similarity)

- Design and contribute to further develop existing WIPO's image search systems. Using state of the art technology, design machine learning algorithms, define data training sets by performing quantitative and qualitative analysis, image analysis and cleaning, studying options to clean the current classifications of images, and identify areas of application for WIPO products and services.
- Offer APIs for: (a) automatic Vienna classification of a trademark image (b) automatic Locarno classification of one or more industrial design image(s) (c) provide descriptors that can be used in a search engine (d) search by image similarities.
- Research new technologies in image classification (using machine learning), evaluate their potential, and contribute to advising the Brand and Design Database section on possible new application ideas for WIPO's Global Database Division.

Speech

- Collect and clean data training sets from different sources, align speech and transcription at word level.
- Study and carefully evaluate best algorithm and parameters for specific languages.
- Combine various multilingual sources and techniques to improve speech-to-translated text.
- Work on a better integration of WIPO speech-to-text with various tools
- Explore and contribute to the development of near real time speech to text in multiple languages
- Research, implement and evaluate the extension and fine-tuning of existing S2T models such as whisper
- Evaluate and work on improving the operational performance of neural models for S2T

3.   Requirements

Education

- Advanced university degree in Information Technology. An advanced university degree in other scientific field plus two years of relevant experience will be acceptable in lieu of an advanced university degree specialised in IT.

Experience

Essential
- Experience in the Information Technology development or research fields.
- Experience in working on high profile Machine Learning projects and / or contributing to significant scientific progress in Machine Learning.

Desirable
- Experience in machine learning (especially using neural networks).
- Experience in image processing/image classification.
- Experience in machine translation and/or Natural Language Processing.
- Experience in speech processing.

Desirable skills
- Deployment strategies: dockerization, cloud servers, version control systems (SVN and/or GIT)
- Statistics: automatic document classification approaches (Neural Networks, SVN, Knn, EM, ANNs, Naive Bayes…)
- Web/Internet technologies: Java, Tomcat, JavaScript, Jquery, Angular.js, Jsf, Lucene (Solr /ElasticSearch), CSS, HTML 5…
- Databases: nosql techniques, Mysql, Oracle…
- Scripting languages: Python, bash, Perl
- Unix: Ubuntu, Red Hat, configuring Unix remote servers (using command line mode)
- Ability to write user guide, administration documentation and reports in English.
- Excellent decision-making and problem-solving skills.
- A proven background in research (scientific publications).
- Search engines: Lucene / ElasticSearch / Solr

Languages

Essential
- Excellent knowledge of written and spoken English or excellent knowledge of written and spoken French and good knowledge of English.

Desirable
- Excellent knowledge of English.
- Knowledge of other UN languages. A working knowledge of other official languages of WIPO (German, Spanish, French, Portuguese, Russian, Arabic, Chinese, Japanese or Korean) would be an advantage.

Job-related competencies

Essential
- Knowledge of machine learning techniques.
- Strong software development skills, preferably in Python / Java or C++.

Desirable

- Excellent communication and interpersonal skills and ability to maintain effective partnerships and working relations in a multicultural environment with sensitivity and respect for diversity.
- Neural Networks frameworks: (Pytorch, TensorFlow, FastAI …).
- Internet technologies including Lucene (Solr / ElasticSearch), JQuery, Angular.js, JavaScript, JAVA, JSF, Tomcat as well as UI design CSS, HTML, HTML 5 and related technologies.
- Experience with Unix (Linux, bash, tomcat, ant…).
- Experience with software development and deployment (version control systems: SVN, GIT …), usage of containerization (Docker, openshift) and cloud servers.
- Knowledge of IP information tools, products and standards.

4.   Organizational Competencies

1. Communicating effectively.
2. Showing team spirit.
3. Demonstrating integrity.
4. Valuing diversity.
5. Producing results.
6. Showing service orientation.
7. Seeing the big picture.
8. Seeking change and innovation.
9. Developing yourself and others.