

**REQUEST FOR INFORMATION
GINS_GDD/2018/64/RFI**

**WORLD INTELLECTUAL PROPERTY ORGANIZATION
(WIPO)**



**DEVELOPMENT AND INTEGRATION OF OPEN-SOURCE
MARIAN IN WIPO TRANSLATE**

TABLE OF CONTENTS

REQUEST FOR INFORMATION	2
1 INTRODUCTION.....	2
1.1 BACKGROUND INFORMATION ABOUT WIPO.....	2
1.2 WIPO TRANSLATE	2
1.3 OBJECTIVES	3
2 RFI PROCEDURE	4
2.1 GENERAL.....	4
2.2 CONFIDENTIALITY	4
2.3 HOW TO DELIVER THE ANSWER.....	4
2.4 CONTACTS	4
2.5 CLOSING DATE	5
3 DISCLAIMER.....	5
3.1 NO REIMBURSEMENT OF COSTS	5
3.2 RIGHT OF CANCELTION OR MODIFICATION.....	5
 ANNEX A	
RFI RESPONSE FORM	1

REQUEST FOR INFORMATION

1 INTRODUCTION

1.1 BACKGROUND INFORMATION ABOUT WIPO

- 1.1.1 WIPO is the global forum for intellectual property services, policy, information and cooperation. We are a self-funding agency of the United Nations, with 191 member states.
- 1.1.2 Our mission is to lead the development of a balanced and effective international intellectual property (IP) system that enables innovation and creativity for the benefit of all.
- 1.1.3 We administer 26 international treaties dealing with different aspects of intellectual property protection including the Patent Cooperation Treaty (PCT) whose mandate is to publish information concerning the protection of intellectual property.
- 1.1.4 Being an international organization, WIPO needs to publish information available in different languages; users may need automatic translation tools to be able to read a document (e.g. a patent application) available in a language they do not master. WIPO also publish official translations where the translator can take advantage of a pre-translated version using machine translation (MT).

1.2 WIPO TRANSLATE

- 1.2.1 In this context, WIPO has invested in building its own tool called “WIPO Translate”, based on open source tools (previously “Moses”¹, recently “Nematus2/Amunmt” and from now “Marian3”).
- 1.2.2 Patent documents can be complex and translating them even more so. WIPO Translate is a world-leading instant translation tool, specially designed for patent documents. WIPO Translate tool is publically available at: <http://www.wipo.int/patentscope/translate>.
- 1.2.3 In WIPO Translate, we give preferentiality to machine learning approaches; we try to create translation models learning from the data. The final goal is always to provide the best MT in production in term of quality (quality should be competitive compared to commercial tools), efficiency (we target quick translations, less than 2 seconds for one sentence) and scalability (we train our translation models on world class data).

¹ www.statmt.org/moses/

² <https://github.com/EdinburghNLP/nematus>

³ <https://marian-nmt.github.io/>

- 1.2.4 WIPO has chosen the open-source Marian as the best technology to train efficiently on huge data, to provide good quality translation in an efficient way (less than 2 seconds on CPU hardware, less than 0.5 seconds on GPU hardware).
- 1.2.5 Various related publications are listed at the end of document: <http://patentscope.wipo.int/translate/wtapta-user-manual-en.pdf>
- 1.2.6 WIPO Translate is free of charge and available through the PATENTSCOPE database

1.3 OBJECTIVES

- 1.3.1 The purpose of this RFI is to obtain preliminary information on the availability of potential companies providing support/maintenance and specific adaptations of the open-source Marian, more specifically offering support for adapting Marian in the WIPO Translate context.
- 1.3.2 The information gathered will be used by WIPO to analyse and understand the current state of the market with a view to launching a tender for such services.
- 1.3.3 The tender objective will consist in: further improve the open-source technology Marian in order to offer more possibilities to WIPO Translate.
- 1.3.4 The project will have the following main objectives:
 - Get a better understanding of the selected technology;
 - Keep WIPO Translate “aligned” with the new development of Marian (therefore working on a better integration of Marian into WIPO Translate);
 - Contribute to further exploratory developments of Marian which will have a positive impact on WIPO Translate.
- 1.3.5 The same information will be gathered from different suppliers and will be used to evaluate whether WIPO’s requirements are aligned with market capabilities, assess the overall feasibility and affordability of the project, and identify firms that may be interested in participating in a potential business opportunity relating to the development and integration of the stated open-source Marian in WIPO Translate.
- 1.3.6 Participation in this RFI is not mandatory. Any prospective participant that declines to respond to this RFI will remain eligible to take part in any eventual sourcing process.
- 1.3.7 In view of the information received, participating suppliers may be invited to take part in pre-solicitation information exchanges. This process will allow WIPO to gather additional information, clarify specific aspects of the responses, and validate the intended procurement strategy.

- 1.3.8 Participation in these information exchanges is not mandatory. The practical details of the sessions and the topics to be discussed will be agreed in with the interested suppliers.

2 RFI PROCEDURE

2.1 GENERAL

- 2.1.1 To answer this RFI, participants must fill in the attached form (ANNEX A – RFI RESPONSE FORM).
- 2.1.2 The contact person specified in paragraph 2.4.1 below is available for any clarifications or assistance in case that is needed.

2.2 CONFIDENTIALITY

- 2.2.1 WIPO may incorporate industry comments and responses to this RFI, in part or in whole, into the future release of a solicitation of offers for related goods and services. RFI participants that include data in their responses that they do not want disclosed to the public for any purpose, or used by WIPO except for internal evaluation purposes, must:

- 2.2.1.1 Mark the title page with the following legend:

This document includes data that shall not be disclosed outside WIPO and shall not be duplicated, used, or disclosed -- in whole or in part -- for any purpose other than for WIPO internal evaluation purposes, unless otherwise expressly authorised by [insert company name]. This restriction does not limit WIPO's right to use information contained in this data if it is obtained from another source without restriction. The data subject to this restriction are contained in sheets [insert numbers or other identification of sheets]; and

- 2.2.1.2 Mark each sheet of data it wishes to restrict with the following legend:

Use or disclosure of data contained on this sheet is subject to the restriction on the title page of this document.

2.3 HOW TO DELIVER THE ANSWER

- 2.3.1 Companies interested in responding to this RFI can submit their response by e-mail to the point of contact indicated below.

2.4 CONTACTS

- 2.4.1 All correspondence and questions relating to this RFI will be addressed to the following point of contact:

Mr. Chris PIEKOSZEWSKI
Email: chris.piekoszewski@wipo.int

2.5 CLOSING DATE

- 2.5.1 Responses to this RFI must be received by the point of contact specified above, not later than **Wednesday, 18th April by 17:00 CET**

3 DISCLAIMER

3.1 NO REIMBURSEMENT OF COSTS

- 3.1.1 THIS IS A REQUEST FOR INFORMATION ONLY AND YOUR RESPONSE SHALL NOT BE CONSTRUED AS AN OFFER. THIS REQUEST FOR INFORMATION DOES NOT COMMIT WIPO TO PAY ANY COSTS INCURRED IN THE PREPARATION OF ANY SUBMISSION OR PARTICIPATION TO THIS REQUEST FOR INFORMATION, OR TO CONTRACT FOR GOODS OR SERVICES.

3.2 RIGHT OF CANCELATION OR MODIFICATION

- 3.2.1 WIPO RESERVES THE RIGHT TO CANCEL, AT ANY TIME, THIS REQUEST FOR INFORMATION PARTIALLY OR IN ITS ENTIRETY. NO LEGAL LIABILITY ON THE PART OF WIPO FOR PAYMENT OF ANY SORT SHALL ARISE AND IN NO EVENT WILL A CAUSE OF ACTION LIE WITH ANY PROSPECTIVE PARTICIPANT FOR THE RECOVERY OF ANY COSTS INCURRED IN CONNECTION WITH THE PREPARATION OF DOCUMENTATION OR PARTICIPATION IN RESPONSE HERETO. ALL EFFORT INITIATED OR UNDERTAKEN BY PROSPECTIVE RFI PARTICIPANTS SHALL BE DONE CONSIDERING AND ACCEPTING THIS FACT.

ANNEX A**RFI RESPONSE FORM**

Question	Answer
Company name	
Company address	
Company web page	
Main products/services	
Main market/customers: i. Private companies: ii. National/regional institutions: iii. International Organizations:	
Number of years in the market	
Company location(s)	
Number of Employees	
Financial information	
Last year turnover	
Last year gross margin	
Last year profit	
Contact person and responsible for answering this RFI	
Telephone	
Email	
Please describe how your company is organized or provide your employees' CVs (you may attach one page) a. What is the profile of your employees in terms of research (academic degrees obtained, scientific publications)? b. Do you employ any external collaborators? If so, how many? What is their profile?	
Since when was your company involved in the development of Marian open source (please provide evidence, e.g. date of the first "commit" by one of the company's employees).	
What is the number of "commits" done by companies' employees in the Marian open-source? i. on yearly basis: ii. on monthly basis: iii. on daily basis:	

Question	Answer
What are the development tools you use (programming languages)?	
<p>Please provide a short description (~ 3 sentences, max 400 characters) explaining how your company would approach each of the following technical challenges:</p> <ol style="list-style-type: none"> a. How would you train a model where we have sparse in-domain data (e.g. 10'000 sentences) and big out-of-domain data (e.g. 8M sentences)? b. How would you implement “constrained beam-search” in the following use cases: <ol style="list-style-type: none"> i. Enforcement of placeholders in translations for the translation of entities like persons, places... ii. “Black-listing” of words in output (force the decoder to avoid the use of a list of words) iii. “White-listing” (enforcement of words that we want to see in the output) iv. “Monotonic” translation (force the decoder to avoid any reordering, typical use case: for transliteration). c. Describe how you would rewrite Marian to be able to decode on CPU (replacing amunmt) d. Describe how you would use (or adapt) Marian to train mutli-language encoder-decoders: <ol style="list-style-type: none"> i. One to many: train with English as source and various languages as target ii. Many to one: train with many languages as source and English as target iii. Many to many: train with many languages as source and many languages as target e. Describe how would you export sentence embeddings (use a bitext corpus to train an encoder that has a similar representation for similar sentences) 	
Any other information which may give WIPO a better insight into your company profile. You may attach a company brochure, if available.	