


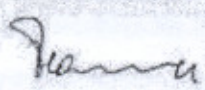

 Implementation Plan for the Language Technology Services initiative of the European Patent Organisation for Romanian			
Contact persons in the EPO		Contact persons in the Romanian State Office for Inventions and Trademarks (OSIM)	
	Richard Flammer, PD 5.4 Tel: 43 1 52126 5400 rflammer@epo.org		Bucura Ionescu bucura.ionescu@osim.ro
	Paul Schwander, D 2722 Tel.: +31(0)70340 2332 pschwander@epo.org		Bogdan Boreschievici bogdan.boreschievici@osim.ro
Common goals: - to provide a machine translation service for Romanian, such that the technical content of documents is understandable to technically qualified users skilled in the art (fit for purpose); - to collect all possible Romanian documents with technical content that have a corresponding English, French or German human translation, or vice versa, in sufficient quality and quantity for a proper training of the machine translation engine.			
Steps to be taken by the EPO:			
Q1 2012	Delivery of lists with patents not yet available at the EPO.		
Q1 2012	Abstracts and titles RO-EN sent by OSIM: check what is available in DOC DB. Ask bulk transfer delivery if needed.		
Q3 2012, Q3 2013, Q3 2014	Corpora repository: Process (OCR, translation memory) and validation of data; transfer to external parties for training.		
Q3 2012	Translation quality: Preparation of IPC class-based test files in order to measure the translation quality.		
Q4 2012, Q4 2013, Q4 2014	Translation training period		
Q1 2013, Q1 2014, Q4 2014	Evaluation of translation quality: by native speakers on basis of prepared test files.		
Steps to be taken by the OSIM:			
Q1 2012	EP validated 14 043 exist. 3672 in XML, 3397 on paper, 6977 pdf. Pdf and xml will be sent (Romanian) version.		
Q1 2012	IPC 7 translated in RO in Word format.		
Q2 2012	Completion: Delivery of documents: EPO needs all Romanian documents in the excel lists (EP validations, National specifications) that have been exchanged between EPO and OSIM. EPO will check and notify which Romanian documents are still missing at EPO. OSIM will transfer these documents (via sftp exchange).		
Q2 2012	Scanning of remaining paper documents		
Q2 2012	3679 PCT in pdf to be sent in Romanian.		

Q2 2012	Compile list of what can be found additionally potentially ca. 10 000 older documents.
Q3 2012	1308 RO patents having family sent to EPO (in PDF).
Q3 2012	360 pipeline protection documents on paper check matching with equivalent in EN GB DE.
Q3 2012	Medical and pesticides official documents and books check those sources with copyright input.
Q3 2012	Translation quality: Contribution in the creation of the IPC class-based test files.
Ongoing until Q4 2013	Delivery of any new NPL/ patent documents with correspondence in EN/DE/FR.
Q1 2013, Q1 2014, Q4 2014	Evaluation of translation quality: Contribution by reading prepared translations and providing quality scores.
Dates of common revision of progress: March 2012, March 2013, March 2014	
	Deputy Director General ALBANO CRISTIAN STRENC 
Signature	Signature

Annex: Description of Language Technology Services programme

Annex to Implementation Plan for LTS

Language Technology Services programme

The programme on European Language Technology Services (LTS) for Patents (CA/29/10 Rev.1), started on 1 November 2010 and will end on 31 October 2014. Its purpose is to provide machine translation services, capable of translating patent documents from English, French and German into all languages of the EPC contracting states, and vice-versa, such that the technical content of documents is understandable to technically qualified users skilled in the art (fit for purpose).

The programme is built upon the concept that the EPO and the National Patent Offices will act as a bridge to the experts, by making available to translation engine providers the necessary patent corpus in order to optimise their machine translation technology for the specific technical language used in patents. The programme consists of four phases, as follows:

- **Phase 1:** building corpora of patent documents - collecting patent documents to enable the creation of a centralised repository of patent corpora in all languages of the EPC contracting states.
- **Phase 2:** language technology services delivery - establishing progressively language technology services for all available languages.
- **Phase 3:** integration - intelligent integration of the language technology services into existing tools and services.
- **Phase 4:** maintenance - securing the sustainability and continuous improvement of the services over time.

Corpora collection for less resourced languages

The phase of the corpora collection is regarded as the most important one. It represents the cornerstone for the set up of the machine translation services, as the corpora is used to train the machine translation engine and its quantity and quality has thus an impact on the translation results.

The corpora generally sought for in LTS Phase 1 includes patent document pairs (e.g. national validations of EP patents and EP patents with national priorities) and national filings. However, a group of languages has been identified for which the available patent corpora may not be sufficient and will thus need additional attention and effort. In these cases alternative corpora sources may have to be taken into account, e.g. scientific journals, university publications, user manuals for products, etc.

The languages in this group are: Albanian, Bulgarian, Croatian, Czech, Estonian, Icelandic, Latvian, Lithuanian, Macedonian, Polish, Romanian, Serbian, Slovak, Slovenian, Turkish.

Scope of this Implementation Plan

This Implementation Plan concerns Phase 1 of the programme, namely the collection of corpora specifically for each of the languages which have been identified as less resourced. It outlines the main actions and deliverables to be carried out and provided respectively by the EPO and the NPOs in order to ensure a successful implementation of the machine translation services for each of these languages.