



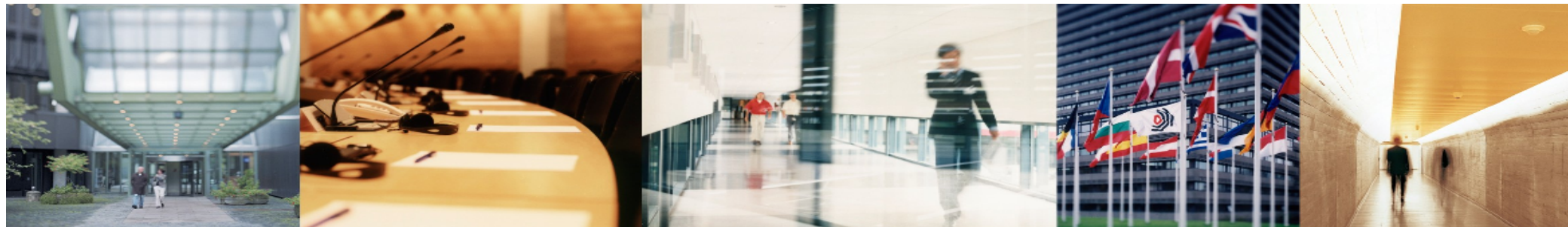
Europäisches
Patentamt
European
Patent Office
Office européen
des brevets

The Challenge of Machine Translation of Patent Specifications and the Approach of the European Patent Office

Georg Artelsmair

Head of Department European Affairs/Member States
European Patent Office

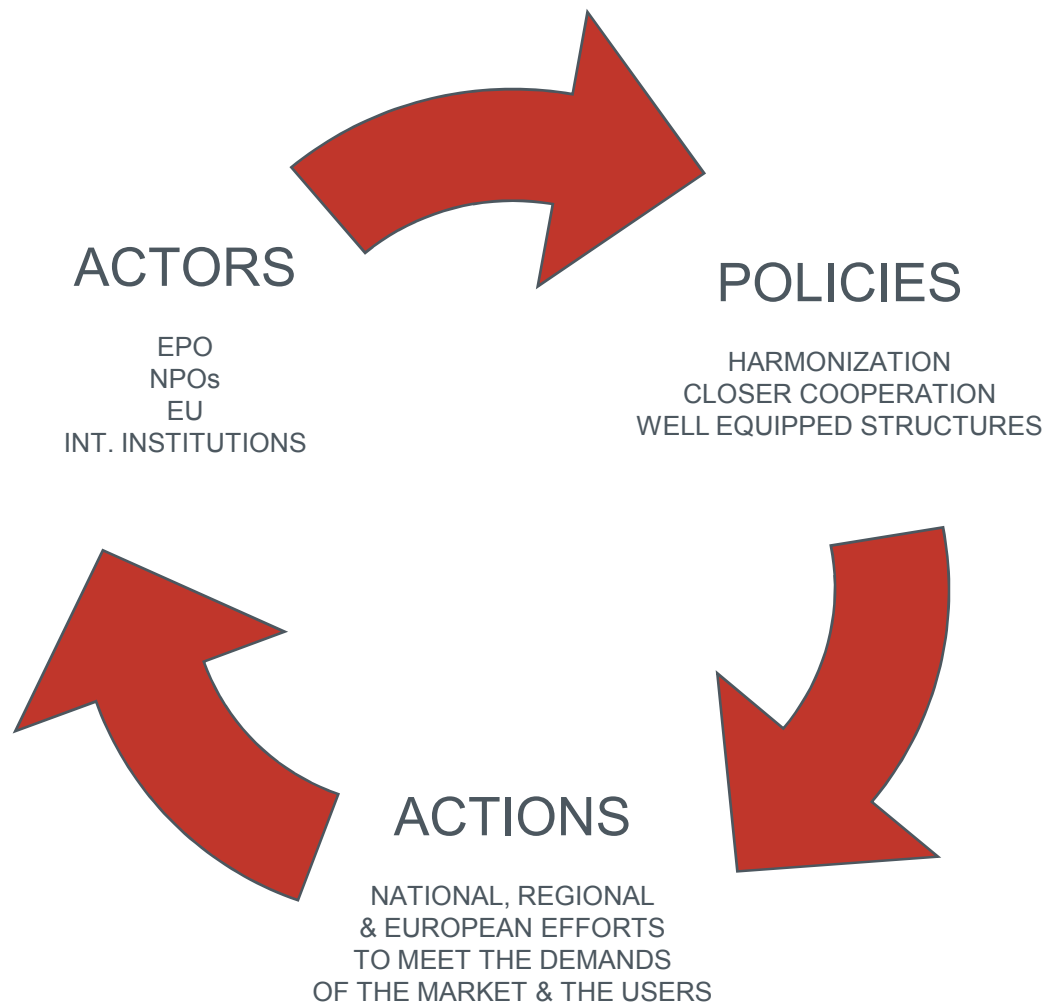
Ottawa, 30 August 2009



Contents

- MT as a co-operation programme in Europe
- European Machine Translation Programme (EMTP): Current approach, status and shortcomings
- Aims of the future enhanced EMTP programme; Aim, justification, challenges and roadmap

Co-operation/ networking in Europe



Current EMTP Approach

- EPO buys a service for offering machine translations. We are not in the business of building translation engines
- IPC classified document pairs are collected together with the national offices involved
- Added value: Developing technical dictionaries based on the IPC classification scheme together with external suppliers
- Current implementations are with English as a pivot language
- Current engines are rule based
- Target group: Technically skilled persons
- Free service
- For patent information only

Status per Language

- **German**
 - Co-operation between the EPO and the German Patent and Trademark Office, the Austrian Patent Office and the Swiss Federal Institute of Intellectual Property
 - Nearly 60,000 translation requests in December 2008
- **Spanish**
 - Co-operation between the EPO and the Spanish Patent and Trademark Office
 - Nearly 35,000 translation requests in December 2008
- **French**
 - Co-operation between EPO and the Belgian, French, Monaco and Swiss IP Offices
 - Operation planned for 4Q 09

Status per Language – IPC Dictionaries in Progress

- **Italian**

- Co-operation between the EPO and the Italian Patent Office
- Provisionally in production with 60,000 entries
- Serious quality problems presently under investigation

- **Swedish**

- Co-operation between the EPO and the Swedish Patent and Registration Office
- Technical dictionaries complete
- Serious quality problems presently under investigation

- **Portuguese**

- The English→Portuguese language direction planned for 2010

Main shortcomings of the present approach

- No suitable rule-based translation engines for certain EPO languages
- English as only pivot language
- Lack of IPC classified document pairs



Need for an enhanced programme

Aims of the future enhanced MT programme at the EPO

Explore available MT technologies that will enable us to introduce a machine translation system in the next 5 to 10 years, which

- ➔ can be tailored for translating patent specifications
- ➔ offers machine translation services from the three EPO official languages into all EPO languages

Will be done together with the national patent offices and European Commission

Justification

- To enable access to patent information to enterprises, researchers and technically qualified users in Europe
- To support the London agreement
- To serve as a contribution to resolving the translation/language issue related to the Community Patent
- To make national documentation available to EPO examiners

Challenges

- Finding a suitable translation engine for languages such as Finnish, Romanian, Estonian, Turkish, etc.
- Managing patent specificity, in particular if there are not enough patent document pairs for a certain language
- Ensuring the quality of translations, in particular given the sometimes low quality of the source text
- Best integration of MT systems into existing and future tools
- Making optimum use of existing human translations (translation memory)
- Operating, maintaining and improving of the MT service

Roadmap

- Workshop with member states and the European Commission in Rome on 26/27 November
- Presentation of a co-operation programme to the Administrative Council in 1st half 2010
- From Q3 2010 onwards collecting document pairs, together with the national patent offices
- Developing the modified engines based on these document pairs
- Presently ongoing projects will be continued

Thank you for your attention

