# DELIVERABLE

| | |
|---|---|
| **Project Acronym:** | **PLUTO** |
| **Grant Agreement number:** | **250416** |
| **Project Title:** | **Patent Language Translations Online** |

## Deliverable 7.1 Survey Structure

**Authors:**

John Tinsley (DCU)
Joeri Van De Walle (Cross Language)
Heidi Depraetere (Cross Language)

| Project co-funded by the European Commission within the ICT Policy Support Programme | | |
|---|---|---|
| **Dissemination Level** | | |
| **P** | **Public** | |
| **C** | **Confidential, only for members of the consortium and the Commission Services** | **x** |

# REVISION HISTORY AND STATEMENT OF ORIGINALITY

## Revision History

| Revision | Date | Author | Organization | Description |
|---|---|---|---|---|
| 1 | 06/04/11 | J. Tinsley/H. Depraetere | DCU/CL | Document creation |
| 2 | 08/04/11 | J. Tinsley | DCU | Added content and structure |
| 3 | 11/03/11 | J. van de Walle | CL | Updated questions |
| 4 | 15/04/11 | J. Tinsley | DCU | Reviewing, copy-editing, formatting |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

## Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

## *Table of Contents*

# Executive Abstract

This deliverable describes the structure of the survey that will be presented to the project consortium's user group partner, WON, to ascertain the adequacy of the PluTO translation engine as relates to the needs of patent users. This will be done by means of a carefully constructed experiments in which we will gain objective feedback from users based on an authentic replication of a specific patent user workflow.

They survey will also comprise a set of more general questions that will help information future directions of development in the project in terms of a number of factors such as language pairs we should address, features to include in the web application and prototype, and user groups we should target for our exploitation plans.

# 1 Introduction

This deliverable document describes our approach to engage the WON user group in assessing two items: a) the translation needs and wants of the patent users, and b) the adequacy of PLuTO translation services in addressing these. In order to put these items in context, we first summarise the various processes followed by IP professionals as relates to patents with particular emphasis on their search and translations activities. We aim to establish a number of details: at what stage search and translations are needed; what is the level of quality required; in what way can PLuTO technologies be exploited to facilitate this; and who exactly uses these services?

# 2 Background: Patent User Workflows

In this section, we give an overview of two of the potential workflows of a patent user (IP specialist, patent search, patent attorney, etc.) in which translation is an integral aspect: patentability and freedom to operate.

## *2.1 Patentability*

One of the goals of the patentabilty process is to establish whether a new invention contains some novelty factor. In order to do this, patent searchers must carry out a prior-art search to collect all existing patents which may be relevant to the new invention in question. Searches are typically carried out on three types of databases/collections: orginial full texts, added value databases, and specific collections (e.g. chemistry collection).

On average, a patentability search returns between 5 and 100 potentially relevant documents, 30% of which will typically require a translation of their abstracts and claims to determine whether they should be considered relevant. The other 70% are documents will be in the language of the original search.

At this stage, perfect translations are not required. In practice, machine translation is often used for languages where there exists a system of sufficient quality. Depending on the quality, this service may be free or paid.

## *2.2 Freedom to Operate*

The process of establishing freedom to operate is needed when an invention reaches the point of testing or commercialisation. At this stage, the user must be certain that another patent is not being infringed upon. This process involves another searching phase that is much more refined than before and is often restricted to a specific regions, e.g. Japan, USA, etc. This search will result in a concise set of potentially relevant documents (on average 5 to 10 documents) on which attorneys must render an opinion.

Depending on the nature of the search, particularly if it is regional, it is likely that most of these documents will be in a single language, e.g. seeking freedom to operate in Brazil will mostly return patents in Portuguese. This will often give rise to a need for translation. At this stage, the translations must be perfect as there are significant legal implications in the case of infringement and users often pay patent agencies for this service. Patent agencies are said to generate up to 65% of their revenue from translation activities.

# 3 User Survey Structure

This section outlines the structure of the survey we will present to the WON user group. The survey is split into two parts: direct questions and an MT adequacy experiment.

In the direct questions section, we will present the users with a number of questions pertaining to their translation needs on a day to day basis and how they currently operate. The aim of these questions is two-fold. Firstly, as we touched upon in Deliverables 3.1 and 6.1, it will serve as a point of reference for ongoing activities in the project as relates to prototype development, such as what languages we should handle and what features we should implement. Secondly, it serves as an information gathering exercise for our exploitation activities in work package 9. The survey will allow us to profile users and identify target groups for our services. This in turn will help to drive the creation of a business model around our software.

The MT experiment relates directly to the adequacy of our translation output for the needs of patent users. Our experimental setup will allow us to get both objective and subjective feedback on the quality of our translation service as relates specifically to the challenge of patent translation. The experiment will be fully developed in conjunction with some members of WON to ensure its authenticity. The outcome of this experiment will inform us whether the translation quality of our MT service (as of year 1) is adequate for users and, if not, what we need to do to improve it.

The profile of the users participating in the experiment will be that of the experienced patent searcher or IP specialist that is a native speaker of the target language. Given the expertise of the WON members, we will focus our experiments on the chemistry, engineering, and electronics domains.

## *3.1 Direct Questions*

- Who are you/what is your profile? (IP specialist, attorney, searcher, other [specify])
- What tools do you currently use? (Espacenet, Patentscope, Google, other [specify])
    - For searching
    - For translating
- What features do you find useful in the tools you use?
- What features do you think are missing from the tools you use?
- Rate the following features from 1 to 5 according to how useful you think they are:
    - translation of copy/pasted text
    - translation from document upload
    - translation from URL
    - translation from patent number
    - integrated post-editing environment
    - downloading translation in desired format
    - browser plug-in to translate search results from any engine
- How do you handle your translation needs currently?
    - I translate the documents myself
    - I use free tools to machine translate documents
    - I use payable tools to machine translate documents
    - I outsource translation
        - Where?
- How much do you spend on translation annually?
- What 3 languages are most important to you in terms of translation needs?
- What section of the patent document is most important to you for translation (in descending order of importance)?

    - title
    - abstract
    - claims
    - description
        - experiment description
- For each section, what level of translation quality is typically required?
    - gisting quality
    - perfect quality
- What potential uses do you see for machine translation?
    - triage: separating relevant from non-relevant documents
    - post-editing: use machine translation output as a basis to speed up translation
    - other...
- Would you be interested in a PluTO run workshop on machine translation and evaluation?

### *3.2 Adequacy Experiment*

Users will be placed in a hypothetical patentability search scenario. Given a description of a brand new invention, users will be presented with 10 potentially relevant documents which have been machine translated. These documents will have been carefully selected to ensure that 50% are relevant to the new invention and the other 50% are not relevant. Users will then be asked for each of the 10 documents whether:

a) it is relevant to the new invention;
b) it is not relevant to the new invention;
c) they cannot tell whether it is relevant given the quality of the translation.

We will then evaluate the answers in terms of precision to assess the adequacy of our system on an objective basis. Additionally, for each document the user will also be asked to make a 'snap judgement' on whether translation of the document is good enough for sorting out the relevant documents. This, in conjunction with the findings of the experiment, will help us gleam a better overall picture of the overall adequacy of our translation service to date.

This experiment will be carried out across 3 IPC domains - chemical, engineering and electronics - as these are the areas of expertise of our user group, WON.

# 4 Summary

Based on initial conversations with our partner WON, we have established the to key areas in the patent user process in which translation is fundamental. Given this information, we have developed a user survey the results of which will provide us with two valuable pieces of feedback: how our current translation systems performs relative to the needs of users, and what direction we should take in our development going forward in terms of language pairs, features, exploitation, etc.

The survey will be carried out between May and August 2011 and the findings will be published in M18 as Deliverable 7.2.