DELIVERABLE

Project Acronym:	**PLuTO**
Grant Agreement number:	**250416**
Project Title:	**Patent Language Translations Online**

Deliverable 7.2 First Report on Survey's Results

Authors:

John Tinsley (DCU)
Joeri Van de Walle (CL)
Heidi Depraetere (CL)

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	Organisation	Description
1	25/08/11	J. Tinsley	DCU	Document created
2	05/09/11	J. Van de Walle	CL	First draft created
3	08/09/11	H. Depraetere	CL	First draft edited
4	13/10/11	J. Van de Walle	CL	Update after status meeting
5	14/10/11	J. Tinsley	DCU	Copy-editing and 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This deliverable contains original unpublished work except where clearly indicated otherwise.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

Deliverable 7.1 submitted at Month 12 of the PLuTO project comprised a survey that was to be taken by members of the user group WON. The aim of the survey was two-fold. Firstly, a set of general questions were asked to establish the landscape in which IP users currently operate in terms of the tools they use for patent searching, for patent translation, and the scale of their operations, amongst other things. The second part of the survey focused on assessing the adequacy of PLuTO MT by means of a usability experiment. This deliverable presents the results from the first part of the survey and offers some analysis of the findings.

For the results of the second part of the survey, the usability experiment, please refer to deliverable D7.6, “First report on the intrinsic and extrinsic quality of MT”.

1. Survey Recap and Motivation

Very often the focus of MT evaluation is on automatic evaluation of the translation quality using metrics such as BLEU, NIST and TER. Although that type of evaluation certainly serves a purpose in the context of the development of the MT systems, it does not necessarily stand in any direct correspondence to potential client satisfaction. With the user survey, we wanted to evaluate exactly that: In how far do end users find the service useful? Are the translations produced by the service of sufficient quality to help patent information specialist with their daily translation-related tasks? Also, the survey provided us with the opportunity to get a better idea of how patent information specialist work today and which tools they currently use for solving translation issues.

The survey was split into two parts, each of which focuses on a different aspect of the service. The first part of the survey asked direct questions relating to how users intend to use the service and relating to features they may or may not find useful, to obtain a better profile of the target users of a potential service, and to establish business opportunities and competition. Examples of questions asked are:[1]

- For which languages do you need translation the most?
- What tools do you use (for searching)?
- What tools do you use (for translating)?
- Please rate the following features according to how useful you think they are

The WON found 84 people willing to take the survey. The findings of this part of the survey are discussed in section 2 "Findings – Direct Questions". For the full list of questions and responses, please refer to Appendix A.

The second part of the survey, which consisted of a test to assess the adequacy of the service, is described in detail in deliverable D7.6, "First report on the intrinsic and extrinsic quality of MT".

2. Findings – Direct Questions

Questions 1 to 6 served to give us an idea of the profile of the users that participated in the survey and how they currently work.

With questions 7, 8, 10, and 11 we tried to get some feedback as to how users thought a translation service might work. Questions 9 and 13, finally, participants were asked to give their opinion on a number of features that a translation service might offer.

In the following sections, we present some of the main findings from these questions.

2.1. User Profile

The majority of participants appeared to work for a multinational (61%) or company with more than one hundred employees (19.1%). Given this, the results reported here may be seen as typically reflective of the operating procedures of larger companies in the IP space.

[1] For the complete list of questions (and answers), please refer to Appendix A.

2.2. *Language Requirements*

Asked about the foreign languages they need translation for the most, 35% percent of the respondents answered that Japanese is the most important language to them. For 32% of the respondents the language they need translation for the most was Chinese.

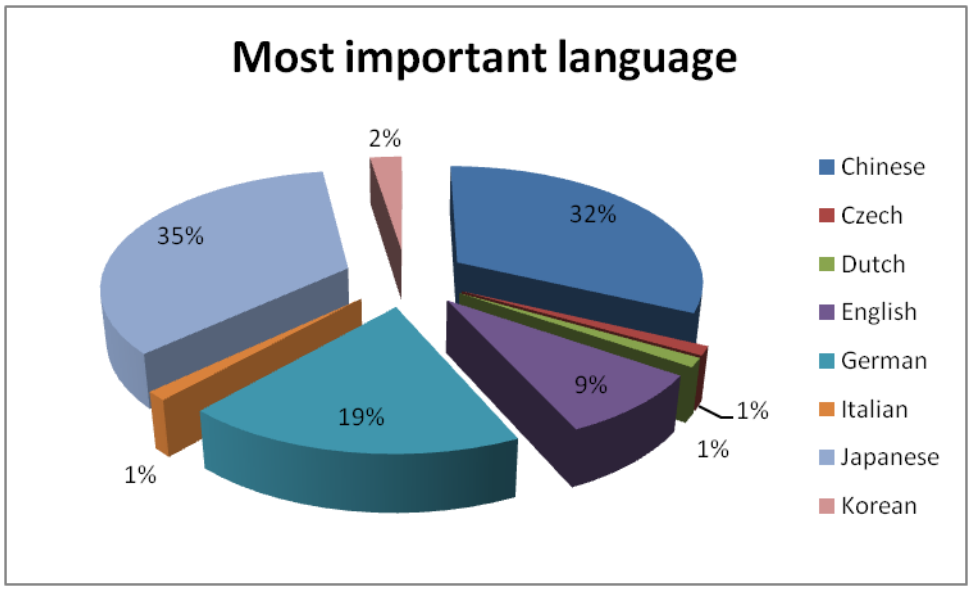

Figure 1: Most important language for translation

For 36% of the respondents Chinese is the second most important language and for 12% it is the third most important language. For 34% Japanese is the second most important language and for 21% it is the third most. Korean, although only seen as the most important language by 2% of the respondents, is the second most important language for 11% of the respondents and the third most important language for 40% of the respondents.

Language	First	Second	Third
Chinese	**32%**	**36%**	**12%**
Czech	1%	0%	0%
Dutch	1%	2%	0%
English	9%	3%	2%
French	0%	6%	1%
German	**19%**	5%	10%
Italian	1%	0%	1%
Japanese	**35%**	**34%**	**21%**
Korean	2%	**11%**	**40%**
Russian	0%	3%	9%
Spanish	0%	0%	4%

Table 1: Most important language summary

Looking at the summary table, it is clear that the most wanted languages are Chinese and Japanese, with Korean and German following in third position. While there is some demand for other languages, certain languages did not feature in our responses at all, e.g. Swedish. While obviously taking into account the profile of those who took the survey, we must consider these findings when selecting and prioritising language pairs going forward.

2.3. Tools

Interestingly enough, 90% of the respondents use free tools to assist them with search-related translation activities, while still 40% turn to external parties to have documents translated by a human translator. Below is an overview of the tools people use.

In anticipation of the tools the PLuTO service might have to be integrated with, the question was asked which tools users currently use when they have to search for patents.

As far as tools (databases) for *search* are concerned, it looks like there is no clear preference for any tool in particular. With 70% percent of the respondents using it, Espacenet seems to be the most popular (free) tool, but also other (paid) tools, such as PatBase, STN, Questel, and Thomson, seem to be used by about half of the respondents (see Figure 2 below). This will influence the integration policy with PLuTO software, e.g. the browser plugin.

3. What tools do you use for searching?

	Response Percent	Response Count
Espacenet	**69.5%**	**57**
Patentscope	22.0%	18
Thomson	39.0%	32
PatBase	53.7%	44
TotalPatent	4.9%	4
Questel	42.7%	35
STN	50.0%	41
Google Patents	17.1%	14
Other (please specify):	7.3%	6
	answered question	**82**
	skipped question	**2**

Figure 2: Search Tools

Asked about the *translation* tools they use, the majority of respondents indicate that they primarily rely on Google Translate or the automated translation services offered by the National Patent Offices for translation related tasks. Each of those services is used by almost 80% of the respondents (see Figure 3 below).

4. What tools do you use for translating?

	Response Percent	Response Count
Dictionary (paper or online)	38.3%	31
Google Translate	**79.0%**	**64**
Yahoo! Babelfish	21.0%	17
Microsoft Bing Translator	1.2%	1
Babylon	2.5%	2
National Patent Office services (JPO, KIPO, SIPO,...)	77.8%	63
Other (please specify):	27.2%	22
	answered question	**81**
	skipped question	**3**

Figure 3: Translation Tools

About 30% of the respondents make use of the tools that are already integrated into their search tools (Espacenet, PatBase, and Questel all offer integrated translation tools).

2.4. PLuTO Service

A first important observation in relation to the PLuTO service is that, although 90% of the respondents currently make use of free translation services, almost two thirds of the respondents (65%) would be willing to pay for a high-quality automated translation service. 60% of the respondents feel a pay-per-use model (where users pay for each document they translate) would be most appropriate in that case.

A large majority of respondents (more than 90%) expect the service to primarily be of use to them for helping them separate relevant from non-relevant documents that are written in foreign languages. Only 30% is convinced that the service may also assist them with other tasks, such as speed up their translation process, translate search query terms, feed translations to other information retrieval processes, etc.

When it comes to sharing corrections they might make to machine translation output, most respondents seem rather hesitant to share those corrections with other users. Just below half of the respondents would not object to sharing their corrections. 17% does not object to sharing corrections with colleagues, but would rather not share them with users outside their organization. The remaining 35% percent would not share corrections at all and would only like to see their corrections used in future translations they make themselves.

2.5. Potential Features

For the majority of respondents (85%), claims are the most important section of a patent when trying to establish relevancy. Consequently, translation of claims is of primary importance.

In the last part of the survey with direct questions, users were given a number of features that they were asked to rate according to how useful they thought they were.

The top 3 features deemed most useful by the respondents are:

Rank	Feature	Description	Rating
1	Translation from patent number by input field	You have the patent number of a patent you would like to translate. You go to the PLuTO web site, paste the number there into a text field and click Translate. The web site returns the translated patent in the browser.	Rated Very useful by 71% of the respondents and Useful by 25%
2	Translation from patent number by right-click (via browser plug-in)	You are searching your favourite patent database. In the result list of your search you see a patent that you want to translate. You select the patent number, right-click, and select Translate. A new tab opens that shows you the translation of the patent whose number you selected.	Rated Very useful by 63% of the respondents and Useful by 33%
3	Translation of current page (via browser plug-in)	You are searching your favourite patent database. You come across a patent that you would like to translate. You click on the patent so its content is displayed. You click on the PLuTO Translate button in your browser toolbar to translate it. A new tab opens that shows you the translation of the patent you were viewing.	Rated Very useful by 58% of the respondents and Useful by 32%

Table 2: Top 3 Features

3. Lessons Learned

The first part of the survey, the direct questions, has revealed that potential users of the PLuTO system seem to agree on quite a number of things.

For one thing, they clearly seem to agree on the languages that are most important to them. It is clear that the languages with the highest demand for translation are the Asian languages, Japanese, Chinese, and to a lesser degree Korean, in particular.

The direct questions have also given us a better view on the tools that patent information specialist suse for search and translation. Virtually everyone uses Espacenet (for search) and Google Translate (for translation). Other popular search tools are PatBase, STN, Questel, and Thomson. With regard to translation, Google Translate has no competition, except from the translation tools offered by the National Patent Offices. It would be interesting to know which Patent Offices are being consulted the most and why. Our guess is that it will be those for which the demand for translation is the highest (and that concern languages for which Google might not perform that well?). So we assume that it is the translation services offered by JPO, KIPO, and SIPO make up for the majority of votes here.

In terms of features, responses have shown that users seem to have a preference for functions that would allow them to translate patents based on patent number. The top 2 features are features that can translate from patent number rather than from patent text. Further investigation and discussions with users will have to reveal more details about the exact uses cases for these features.

Arguably the most important lesson learned from the first part of the survey, for the consortium members at least, is that a majority of users seem to be willing to pay for using a translation service such as PLuTO, provided it delivers on its promise of providing automated translations of high-quality.

All of these findings have been taken into account during discussions leading up to the preparation of Deliverable 9.1 Exploitation Plan. It is clear that there is indeed a market opportunity for a patent translation solution, and we must pay careful attention and consideration to the needs of users when designing and implementing our business model. Please consult D9.1 for more information on this aspect of the project.

4. Action Points and Conclusions

The survey has proven to be of great value. WON has been very active getting their members interested in and enthusiastic about the PLuTO project and the survey response proves that their efforts have paid off. Considering the number of responses, we are confident that the survey is a reliable indicator of the needs and wishes that live among the patent information specialists.

We see direct action points in the following areas of the project:

- Language offering
- Tool support
- Feature support
- Service definition

A first element that requires attention is the set of language pairs we propose to cover. From the survey responses it has become clear that the languages that are currently on the roadmap of the PLuTO project (DE, FR, ES, RU, ES, NL → EN) are not all equally in demand. The need for translating from Dutch, Spanish, or Russian, for example, is a lot less than that for translating from Japanese, Chinese, or Korean. To that end, we have decided to begin the process of data acquisition for the additional Asian languages. As a deliverable in year 2, we will present MT and TMs for English<->German, and Spanish. Additionally, we will develop a prototype Japanese MT engine, as data is immediately available for the English—Japanese language pair.

Thanks to the survey, we now also have a good idea of the search tools patent information specialist tend to use the most when looking for patents (Espacenet, PatBase, STN, Questel, and Thomson). With the exception of STN, which only links to patents stored in other databases, these organisations may be good candidates for discussing opportunities in the context of the PLuTO exploitation. To this end, we have already engaged with PatBase with regards to integration of the browser plugin into their tool.

Finally, with the survey we have collected proof that a majority of potential users are interested in a payable good quality automated translation service and seem to have a preference for a pay-per-use model. This input will be important in the context of defining exploitation service and pricing models.

Appendix A: Survey Questions & Responses

See supplementary document.