Invited Talk 3

Future Plans of Machine Translation System in the JPO

Mr. Eiichi Yamamoto
Patent Information Policy Planning Office
Japan Patent Office

Future Plans of Machine Translation System in the JPO

September 2011, Xiamen

Eiichi Yamamoto
Patent Information Policy Planning Office
Japan Patent Office

Contents

1. Outline for Dissemination Policy of Industrial Property Information in English

3/28

History of Industrial Property Information Web Service

Apr.1996	JPO Official Website launched
Apr.1997	"Patent Abstracts of Japan (PAJ)" retrieval service released
Mar.1999	"Industrial Property Digital Library (IPDL)" started on the website
Mar.2000	Patent Gazette "Full-text Translation" service released
Feb.2003	The Official Website renewed
Mar.2004	Design Gazette "Full-text Translation" service released
Oct.2004	IPDL transferred to National Center for Industrial Property Information and Training (INPIT) Website
Oct.2004	"Advanced Industrial Property Network (AIPN)" released to IP Offices

4/28

Industrial Property Information Dissemination

JPO : Policy Making on Industrial Property Information Dissemination

INPIT : Provision of Basic and Primary Information

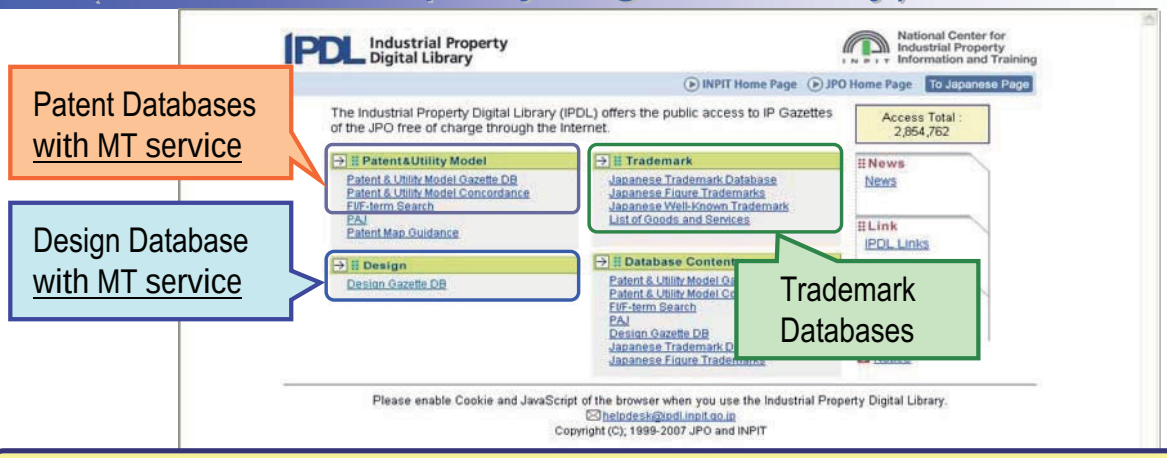
Databases with Machine Translation

IPDL (Industrial Property Digital Library)
Offering 80 million IP information in Japanese and MT English to Public users.

AIPN (Advanced Industrial Property Network)
Offering patent application information and dossiers in MT English to foreign IP Offices.

5/28

IPDL (Industrial Property Digital Library)



English Databases on IPDL

1. Patent Databases

DB with Machine Translation Service
- Patent & Utility Model Gazette DB
- Patent & Utility Model Concordance
- FI/F-term Search

Human Translation Database
- Patent Abstracts of Japan (PAJ)

2. Design Database

DB with Machine Translation Service
- Design Gazette DB

3. Trademark Database

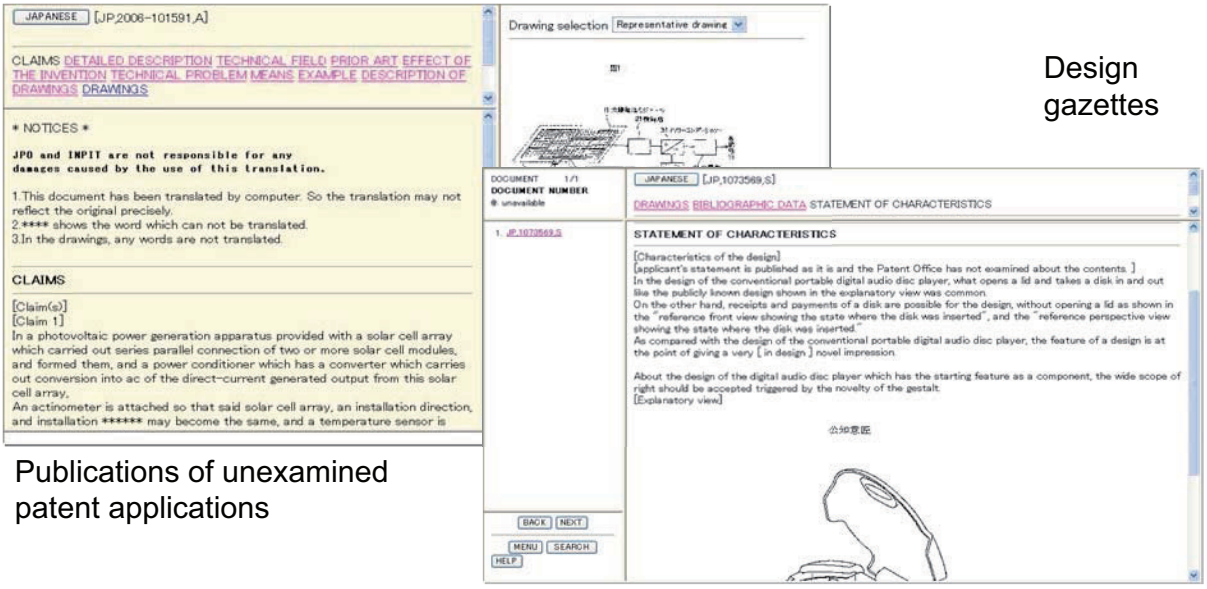
Database (Bibliographic Items such as Dates and Numbers)
- Japanese Trademark Database
- Japanese Figure Trademarks
- Japanese Well-Known Trademark
- List of Goods and Services

6/28

IPDL (Industrial Property Digital Library)

- Full-text Patent and Design Publication by MT -

・Patents and design gazettes can be searched number-searched.
・All information (except drawings) is translated by machine translation system.



Publications of unexamined patent applications

Design gazettes

7/28

2. Mutual Use of Examination-related Information with Foreign IP Offices by Utilizing the Machine Translation

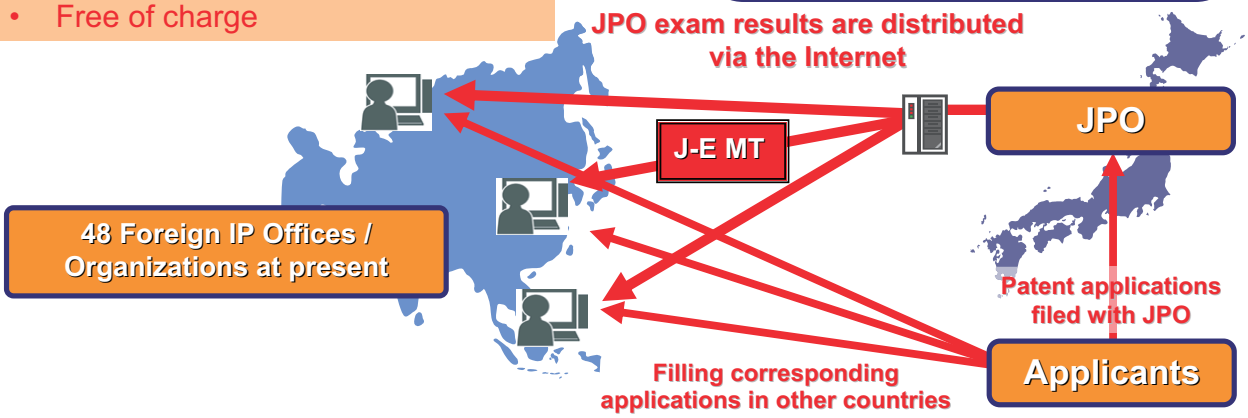
8/28

Provision of the JPO Examination Results in English

AIPN: Advanced Industrial Property Network

AIPN
(Advanced Industrial Property Network)
- Japanese examination-related information is available in English by means of Japanese-English machine translation
- Available for foreign IP Offices
- Free of charge

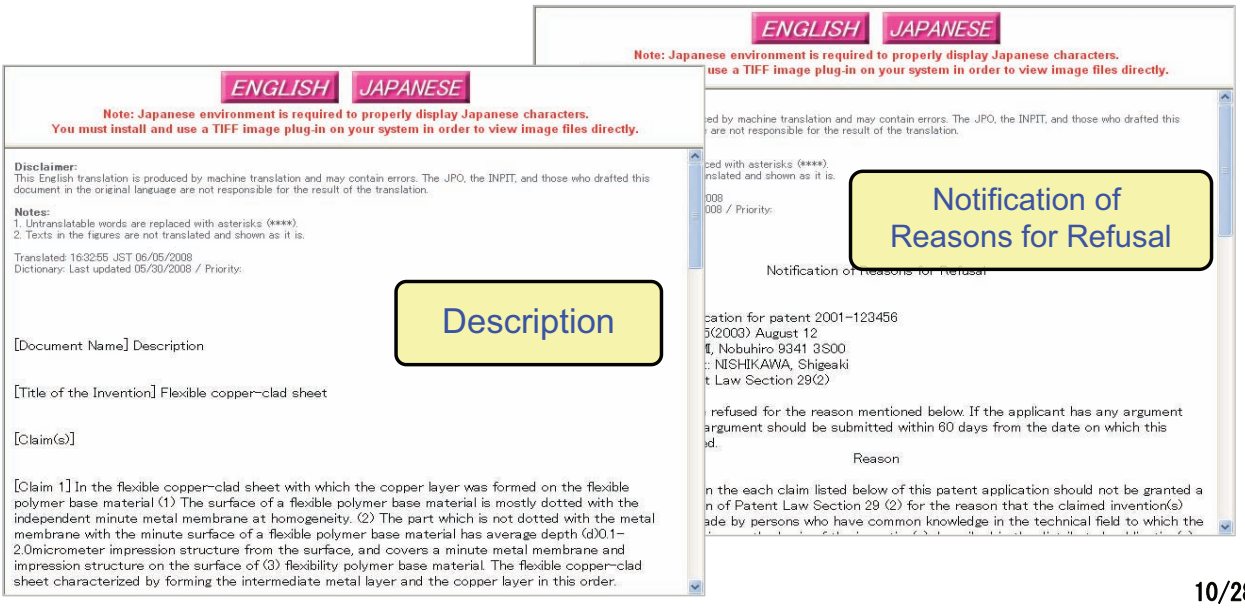
Advantageous Effects
- **Reducing burden on examiners in foreign IP offices**
- **Accelerating JP applicants' acquisition of rights in other countries**



9/28

Advanced Industrial Property Network (AIPN)
- Example of File Wrapper Document Retrieval -

- File wrapper documents are provided in English
- All the Japanese information in FW is translated into English by MT system



10/28

Advanced Industrial Property Network (AIPN)

- Example of File Wrapper Search -

File Wrapper Document List

- The result of file wrapper search comes out in English

Application Info. & Legal Status Info.

File Wrapper Document List

Office Action, Search Report, Written Opinion, Examiner's Note etc.

Patent application number 2001-123456

File Wrapper Document List – Click document title to display contents.
The dates in the list are shown in the format of "dd.mm.yyyy"

- Application data and Transaction history

1. 20.04.2001: Request for a Patent
2. 20.04.2001: Description
3. 20.04.2001: Drawings
4. 20.04.2001: Abstract
5. 02.10.2001: Ex Officio Correction of Defect of Abstract
6. 23.04.2002: Request for Examination (Others' Request)
   -- . -- . ---- : -- Certified or Appended Information
7. 28.05.2002: Written Notice (others' request)
8. 22.07.2003: Search Report by Designated Searching Authority
9. 12.08.2003: Assessment on Search Report by Designated Searching Authority
10. 02.09.2003: Notification of Reasons for Refusal
11. 04.11.2003: Written Argument
12. 04.11.2003: Written Amendment
13. 21.04.2004: Decision to Grant a Patent

[Translation done.]

Examples of document titles in Japanese
(Provided here for reference only)

- 出願マスタ基本項目

1. 20.04.2001:特許願
2. 20.04.2001:明細書
3. 20.04.2001:図面
4. 20.04.2001:要約書
5. 02.10.2001:要約不備職権訂正
6. 23.04.2002:出願審査請求書(他人)
   --.--.----:認定・付加情報
7. 28.05.2002:通知書(他人請求)
8. 22.07.2003:検索報告書
9. 12.08.2003:検索外注利用状況票
10. 02.09.2003:拒絶理由通知書
11. 04.11.2003:意見書
12. 04.11.2003:手続補正書
13. 21.04.2004:特許査定

11/28

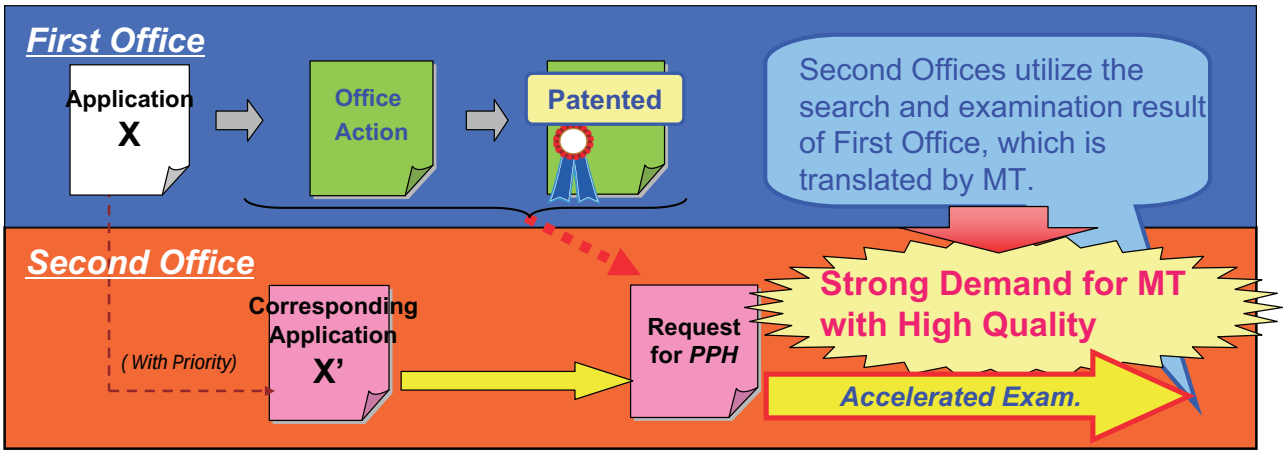
Advanced Industrial Property Network (AIPN)

- Patent Examination Cooperation -

Patent Prosecution Highway (PPH)

On condition that a patent application has been determined to be allowable in the first office, the application is qualified for accelerated examination through a simple procedure according to a request from an applicant in the second office.

- It assists applicants to obtain a foreign patent expeditiously.
- It reduces burden of examination and enhances quality of examination for Offices.



12/28

3. Efforts to Improve Accuracy of the Machine Translation

Efforts to Improve Accuracy of the Machine Translation

- **Collecting, Registering Unknown Terms and Words**

  To collect words and terms which failed to be translated in AIPN or IPDL, and register the corresponding translations in the dictionary (5,000/year)

- **Receiving Report and Feedback from Other IP Offices Overseas**

  Mistranslation Report

  With reports from users (examiners in IPOs overseas including EPO, USPTO) on mistranslations detected in AIPN, the reported mistranslations are reviewed, and if necessary, the appropriate translations will be registered in the dictionary.

  Feedback from the Users

  Fill-in questionnaire is provided on AIPN, to collect information from the users on how they have utilized the JPO examination results.

- **Preprocessing the Documents in Original Language before MT**

  Pre-editing or pre-formatting of document data before MT, replacing with appropriate characters (e.g., changing tags)

Efforts to Improve Accuracy of the Machine Translation

- Enhancement of the memory and pattern dictionaries

Studies for Improvement of Translation Accuracy（J-E machine translation in AIPN ）

Analyzing Notifications of Reasons for Refusal and registering expressions that can be used as fixed sentences in the dictionaries

FY2003: Analyze unknown words, phrases and fixed sentences in Notifications of Reasons for Refusal
→Conclusion: Entering above in translation engine is an effective measure

FY2007: Analyze and extract frequently used unknown terms, phrases and fixed sentences from one of the frequently used provisions (Patent Law Article 29-2) *20,000 Notifications of reasons for Refusal

FY2008: Analyze and extract unknown terms, phrases and fixed sentences from some of the frequently used provisions (Patent Law Article 29(1)iii, Article 29-2 and Article 36) to expand the survey results of the previous year *40,000 Notifications of Reasons for Refusal

Registration of Routine Expressions from Examiners' Drafts (FY2010)

Routine expressions which examiners use in drafts (e.g. notice of refusal, dismissal of amendment) to be registered in the dictionary.

Efforts to Improve Accuracy of the Machine Translation

- Studies for Improvement for Accuracy of the Machine Translation (FY2010)

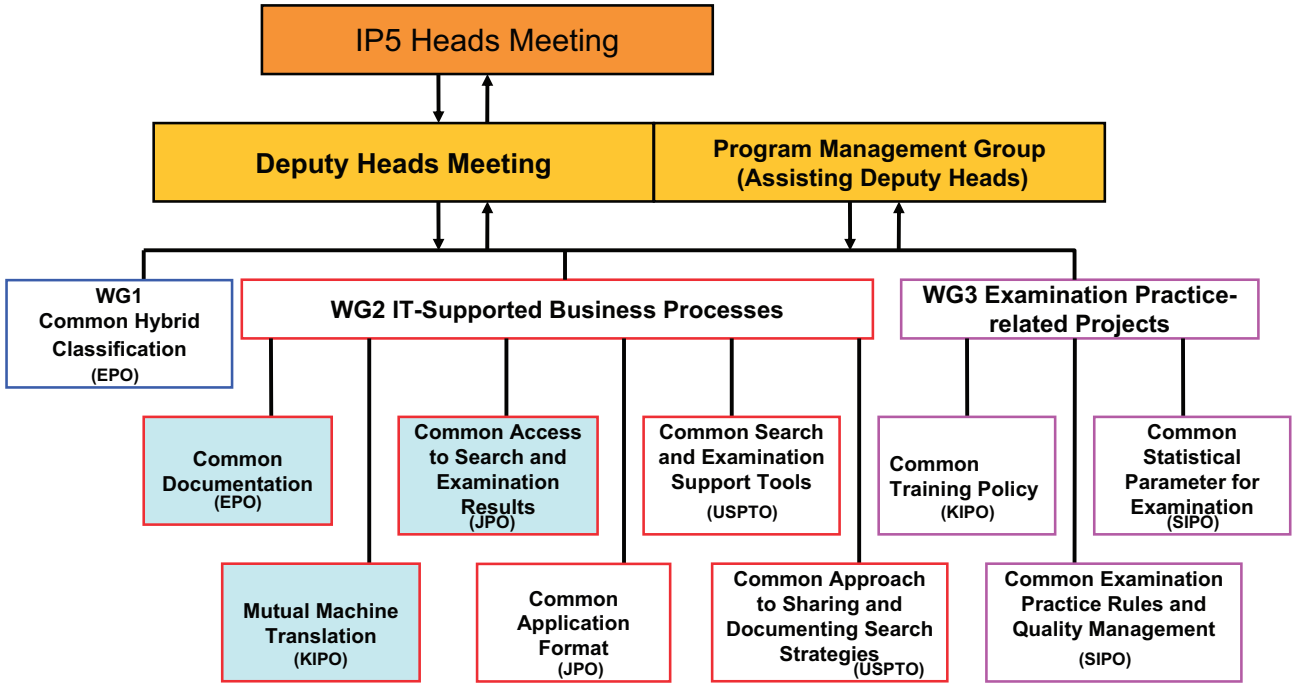
Study on Japanese-Chinese Machine Translation
・Chinese machine translation of Japanese patent applications
・Analysis of accuracy of the Machine Translation and recognition and measures for the challenges.

Study on Japanese-English Machine Translation
・Assessing the quality of Japanese-English machine translation on AIPN
・By conducting the human evaluation and multiple automatic evaluations, examining the accuracy of automatic evaluations

4. Future Vision of the Machine Translation Utilization Conducted by IP5

17/28

IP5 Meetings and Foundation Projects



*( ) stands for the Office in charge of each project.

18/28

Mutual Machine Translation

- Understanding of the Current Quality and the Quality Goal -

Aim: "To help the offices overcome the language barrier of patent information and allow greater access to each other's patent information."

Quality Assessment of Machine Translation

- Non-English speaking Offices (i.e., JPO, KIPO, and SIPO) submitted sample sentences by extracting and machine-translating. English-speaking Offices (i.e., EPO and USPTO) assessed samples by IP5 agreed 5-point scale.

→ Understanding the current translation quality non-English speaking Offices provide

Definition of the "Fit for purpose" Machine Translation Quality

- The target quality level was set to fulfill each purpose.

1) The level of quality which is suitable for prior art searching (score 3)
2) The level of quality which is suitable for substantive examination (score 4.5)
3) The level of quality which is suitable for post grant use (score 5)

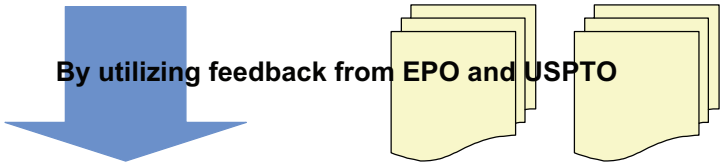
→ Clarification of the target quality of machine translation under the framework of IP5 work-sharing

Mutual Machine Translation

- Improvement of the MT Service by IP5 cooperation -

Error Review and Feedback (September to December, 2011)

As to English machine-translated sentences submitted by non-English-speaking Offices, English-speaking Offices will feed back concretely, based on the criteria such as word selection, phrase order and grammar.



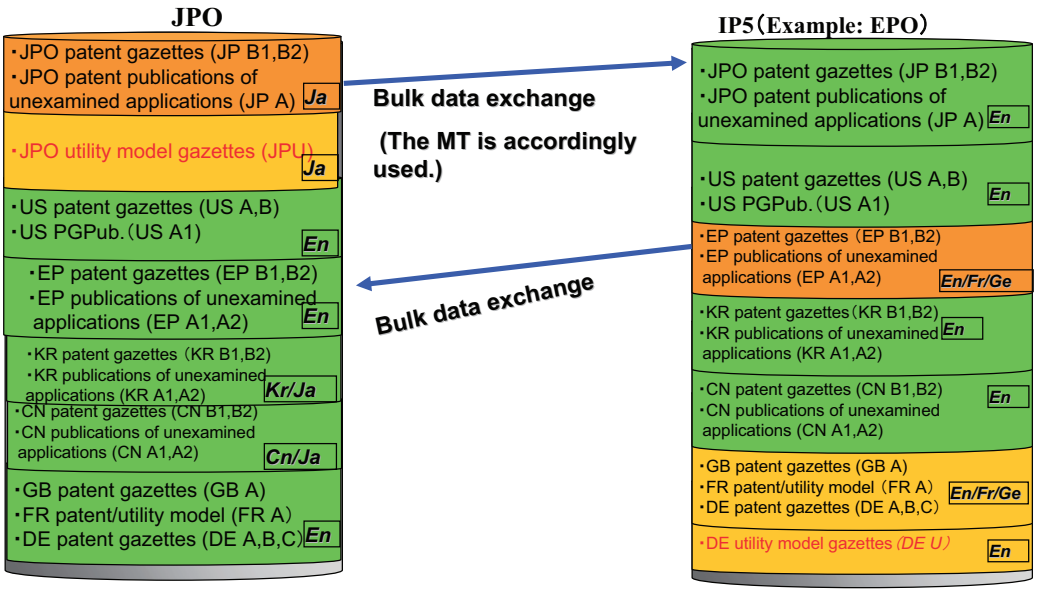
Upgrade of the Machine Translation System in Each Office (January, 2012)

Based on the feedback, non-English speaking Offices will upgrade the machine translation system to improve the accuracy (From January, 2012).

Common Documentation Database

- Establishment and Improvement of In-house Database -

- **In-house databases to be constructed by collecting document types which each Office frequently cites.**
- **The current data exchange scheme to be improved. (Example: standardization of data exchange, medialess exchange, etc.)**
- **The use of text data and the machine translation enable users to use easy-to-read languages as much as possible.**

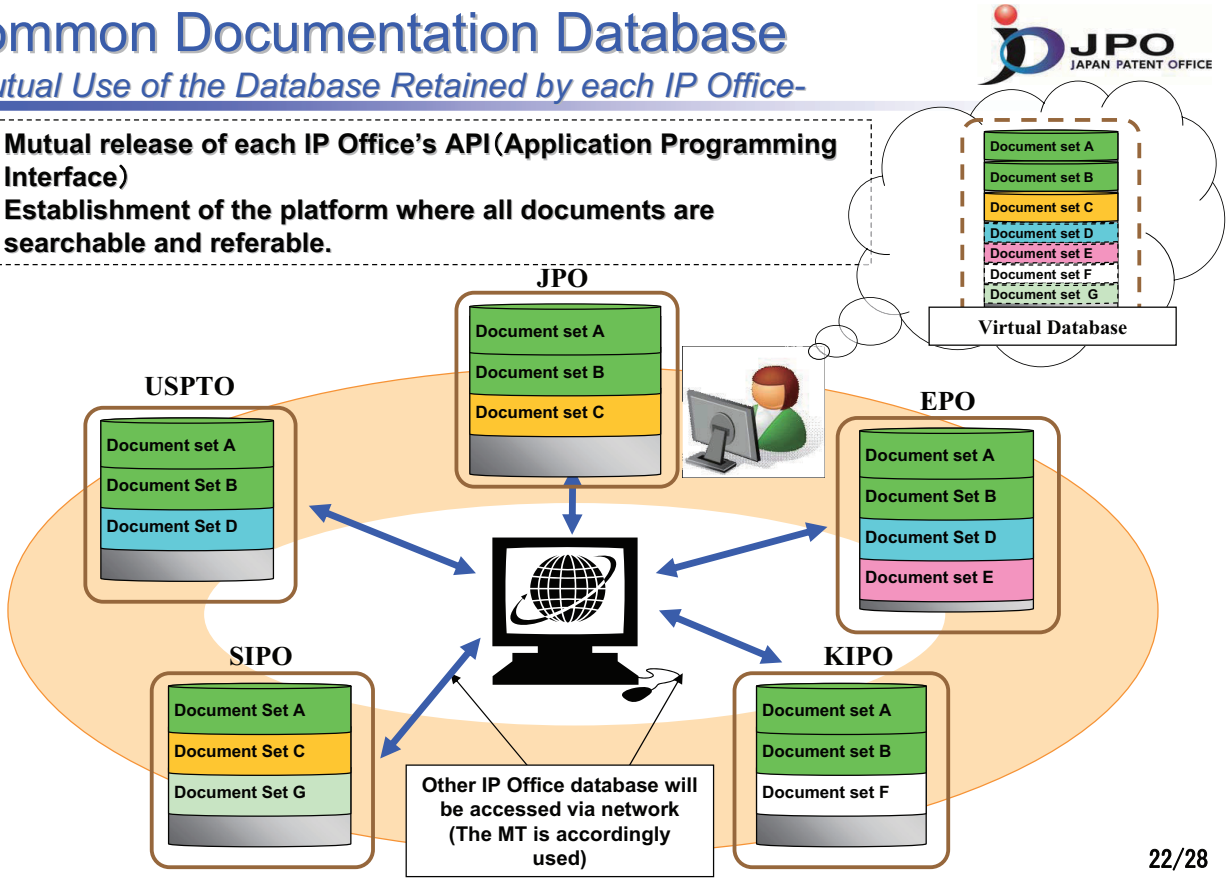


21/28

Common Documentation Database

- Mutual Use of the Database Retained by each IP Office-

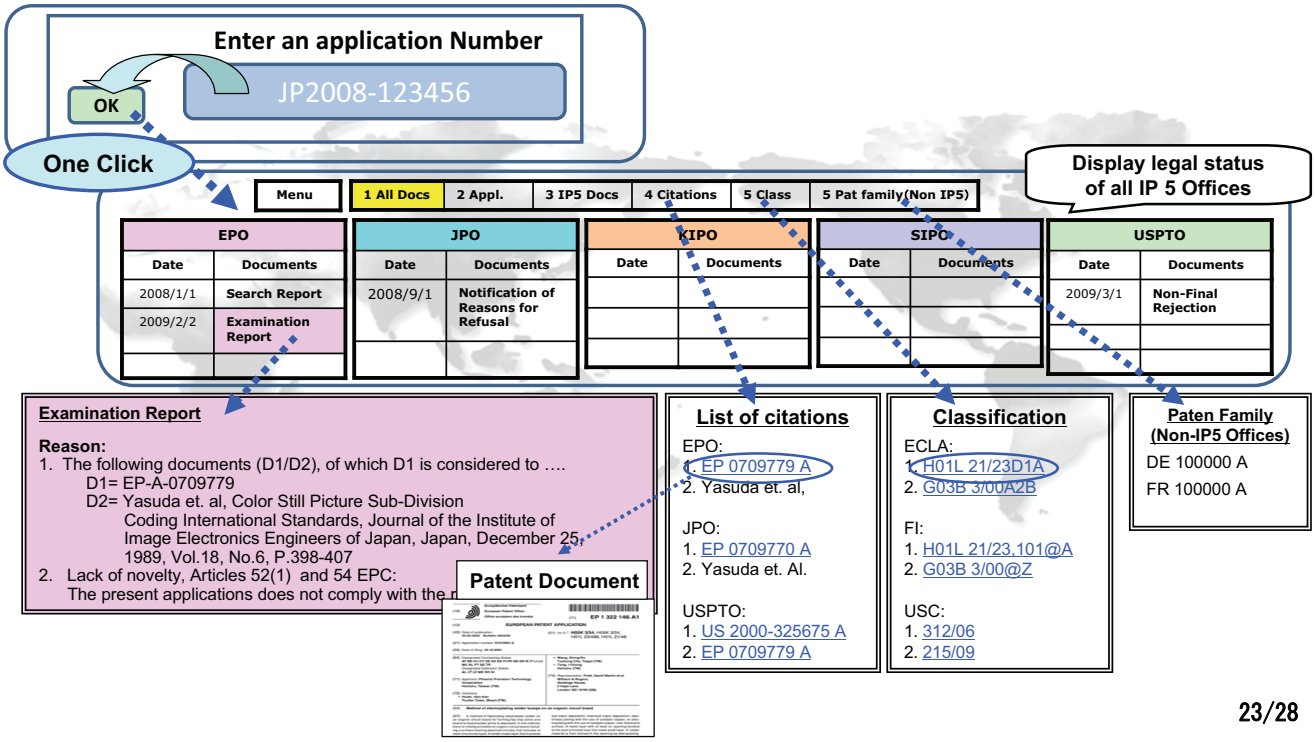
- **Mutual release of each IP Office's API(Application Programming Interface)**
- **Establishment of the platform where all documents are searchable and referable.**



22/28

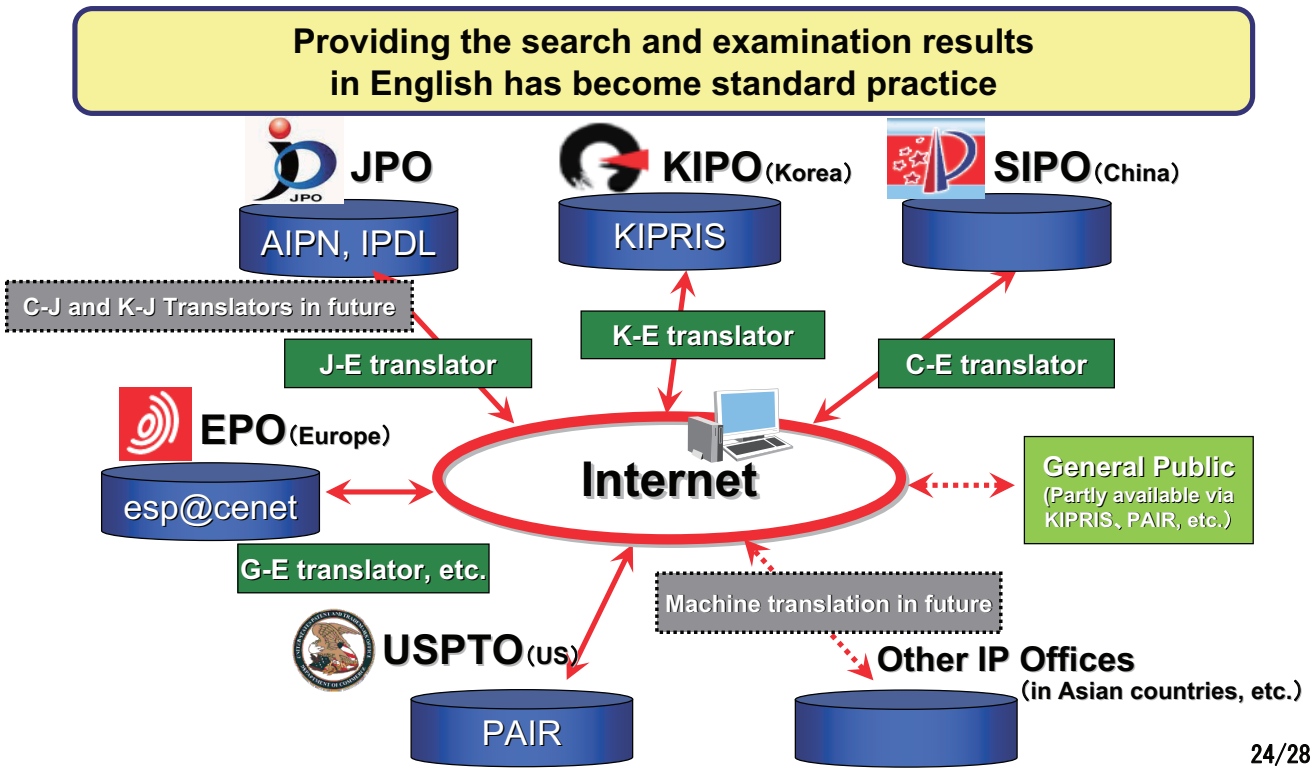
Common Search and Access to Examination Results

- One Portal Dossier (OPD) -



23/28

Machine Translation in Various IP Offices Worldwide



24/28

5. Future Plan of the JPO Operations and Systems, and Further Utilization of the Machine Translation

25/28

The JPO's New Search System

Goal:

- Constructing the state-of-the art IT environment, which enables the leading, speedy and accurate examination.
- Establishing the environment of patent information utilization which contributes more greatly to R&D and the management strategy of companies and universities.

- Introduction of the following search function and system to be considered.

- Cross-lingual search system (Chinese, Korean, etc.)
- Concept search system, etc.

26/28

Research on Cross-language information retrieval

The JPO conducted research on the development of the next-generation search system regarding cross-language information retrieval system.

- Evaluation of the effectiveness of the two translation methods (Keyword translation, Contents translation)
- Evaluation of translation results (Korean-Japanese, Chinese-Japanese)

Korean-Japanese
→Translated sentences are fairly understandable. (practicable level)
Chinese-Japanese
→Further development is needed. (especially, enhancement of the dictionaries)

Types of Dictionary	Chinese	English	Remarks
Basic Vocabulary	250 thousand words	1.03 million words	Chinese are 1/4th of English words
Technical Term	238 thousand words	2.72 million words	Chinese are 1/10th of English words
Automatically registered words	1300 words	–	–

Thank you for your attention!