Infrastructure for Digitalization of Dossier Information in Vietnam

Phan Phung Tuan
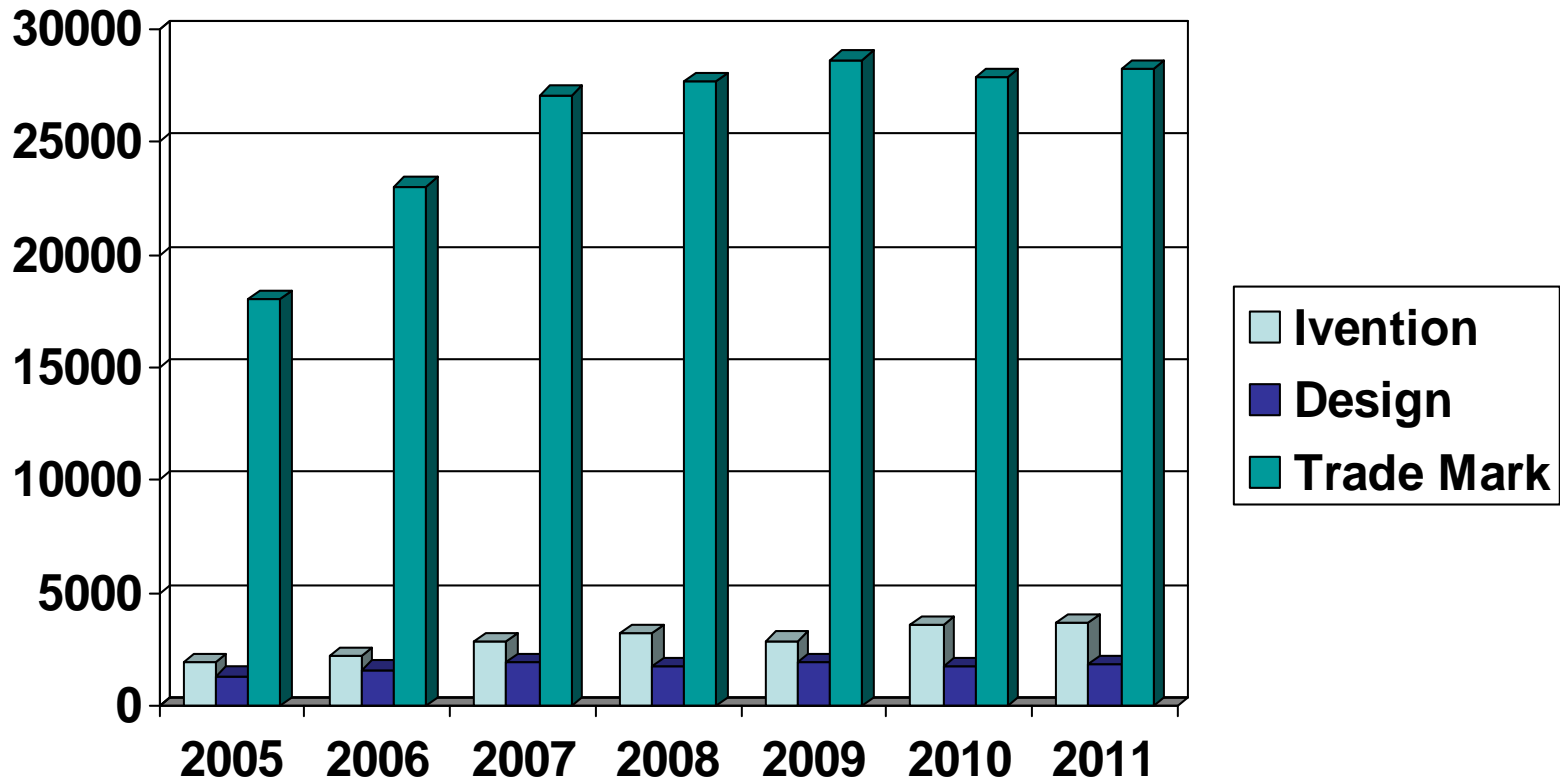
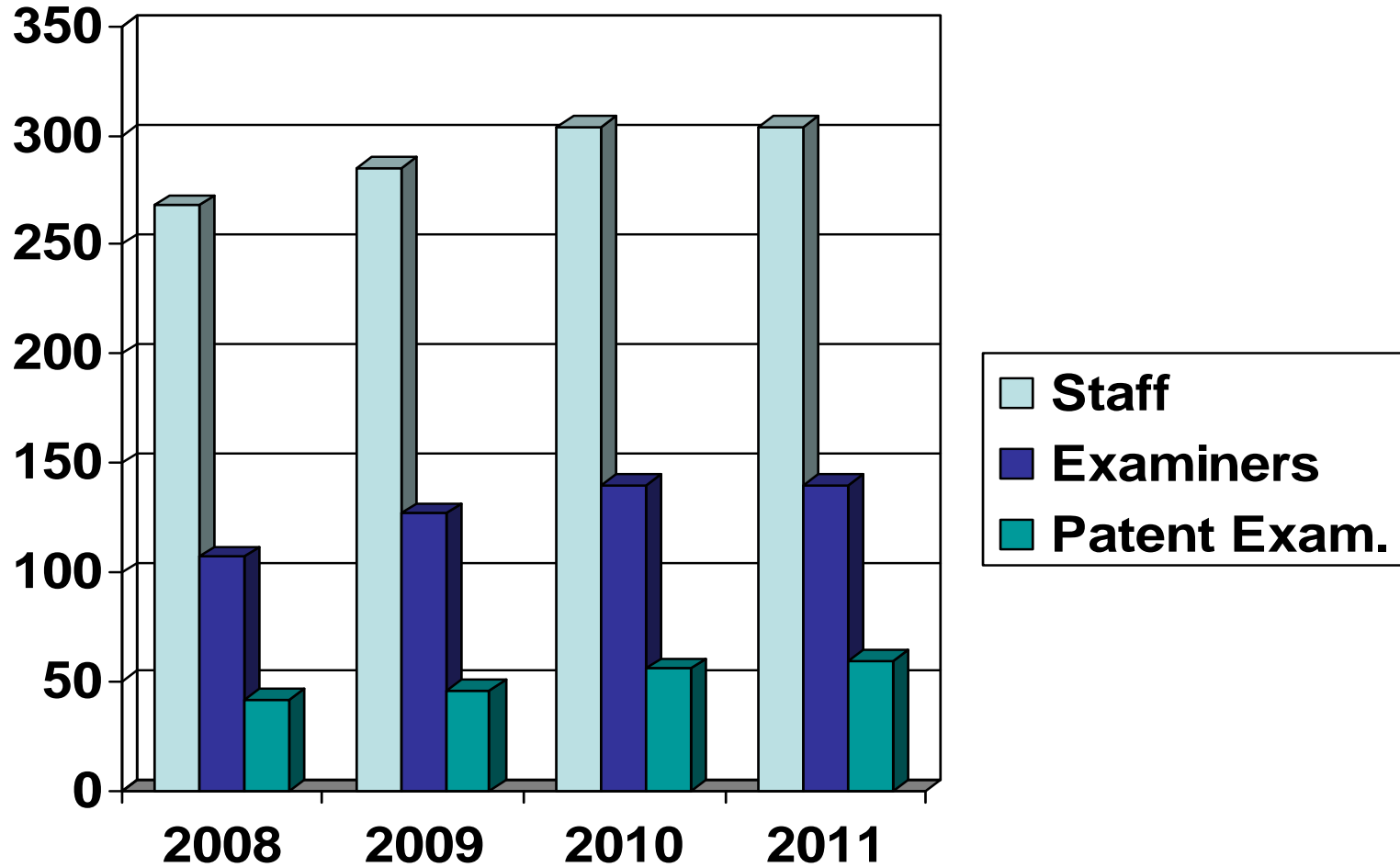NOIP, Vietnam

E-mail: phanphungtuan@noip.gov.vn

Content

1. Introduction to NOIP: The application number, NOIP manpower
2. The Modernization Projects
3. MOIPA Project
4. WIPO Digitalization Project
5. Annex

The application number to NOIP
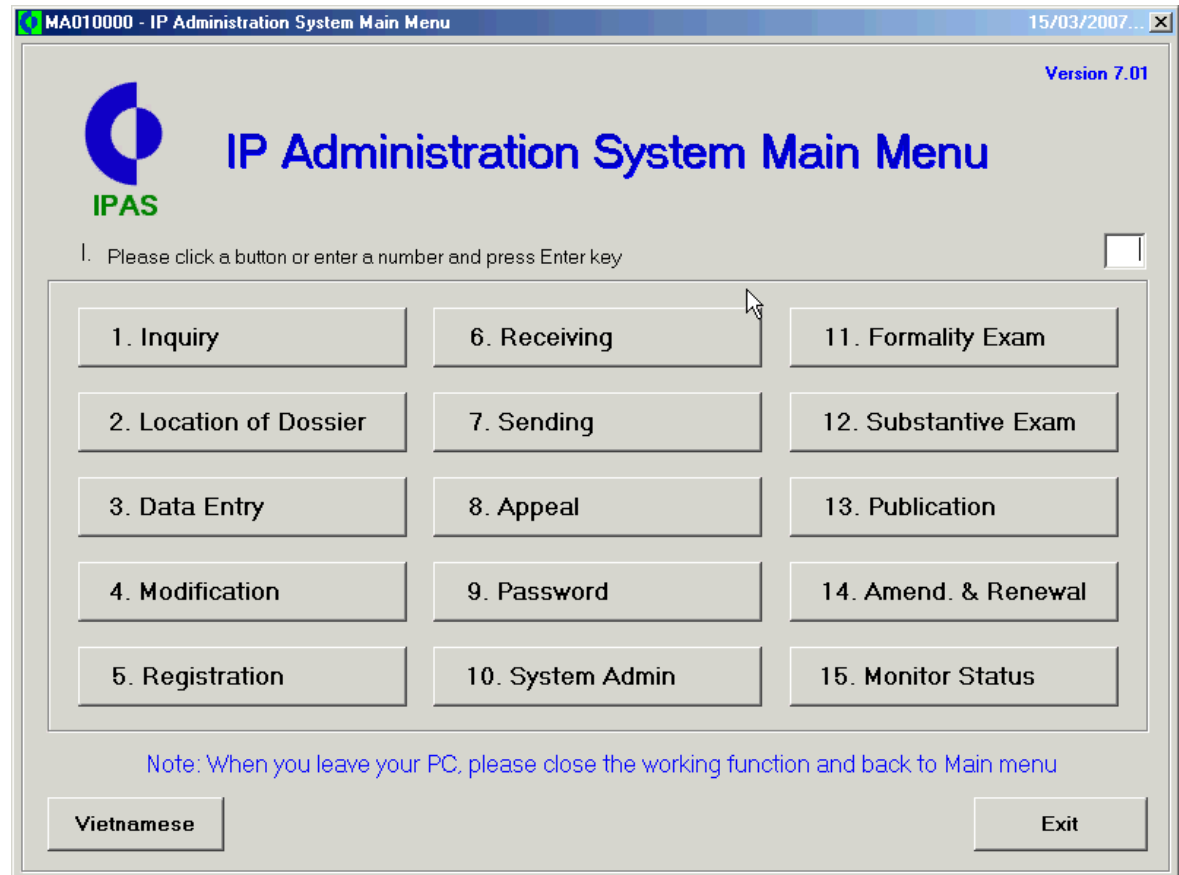
NOIP manpower

2. The Modernization Projects

- From 2001: Japan-Vietnam Cooperation Project start up, IPAS was set up, Data entry function started
- From 2008: WIPO-NOIP cooperation Project started, Digitalization of patent documents from 1982.

3. MOIPA Project

- Time: from 2000 to 2004,
- Cooperation party: JPO
- Out put: Industrial Property Administration System (IPAS)
- Function: Data Entry, Examination, Publication, Registration…

Data Entry Function of IPAS

- Scan application documents (dossier),
- OCR to text for bibliographic information
- (See video)

4. WIPO Project

- Time: 2010
- Project name: "Developing a Software Tool for Digitalization of Patent Documents" of the NOIP
- Use WIPO XML standards for documents and data exchanges

Digitalization project

- Resource: Description for patent from 1982
- Scan in document
- OCR to text
- Review, check
- Publication on XML, can search PDF file

Patent Document

Scan and OCR

Review screen

Thank you
Question?

Annex 1: Scanning system

- The Scanning system shall:
- Take a input a list of documents to scan for gazette;
- Allow for scanning of their pages in TIFF black and white 300 dpi TIFF IV;
- Allow for the quality check of the scanned documents (rescanned bad pages, reordered pages, rotated pages scanned in the wrong directions, support rctor/verso scanning if required, deskew, remove dots);
- Allow for the indexation of the section of the scanned documents (front-page, description, claims, drawings, abstracts);
- Merge the newly scanned documents with the one extracted from IPAS and allow for the completeness check of a gazette;
- Allow for the export of all scanned documents of a gazette, together with bibliographic data and indexation in XML/TIFF ST36 format (detailed format to be specified).

Annex 2: Digitalizing system

- The digitalizing system shall:
- OCR in batch mode all the documents of a gazette, exported from the scanning system;
- Allow for the proofreading of the OCR of the document, if possible indicating to the operators the worse pages and documents to check/correct;
- Export the raw and detailed OCR information for each document in an XML format reusable by the PATENTSCOPE® system (detailed format to be specified);
- Export the text information for each document in an ST36-XML format, as well as in mixed-mode PDF (detailed for amt to be specified);
- Burn the results on optical media (DVD) so that the published documents can be sent to external parties (detailed format to be specified).