

## Committee on WIPO Standards (CWS)

**Second Session**  
**Geneva, April 30 to May 4, 2012**

### SURVEY ON THE IMPLEMENTATION AND PROMOTION OF WIPO STANDARD ST.22

*prepared by the Secretariat*

1. The former Standards and Documentation Working Group (SDWG), at its eleventh session, held in October 2009, approved a questionnaire to carry out a “Survey on the Implementation and Promotion of WIPO Standard ST.22”. (See paragraphs 49 and 50 of the SCIT/SDWG/11/14 and paragraph 52, Task No. 37, of document CWS/1/10 Prov.)
2. According to the decision by the SDWG (see paragraph 51 of the Survey) the Survey was conducted in 2011. The International Bureau of WIPO prepared the questionnaire and made it available in the WIPO Standards Office whereof the Offices were informed by the Secretariat.
3. Thirty Industrial Property Offices took part in the survey, providing their responses to the questionnaire. Received feedback was summarized by the Secretariat in the form of the Survey report which is reproduced in the Annex to the present document.
4. Materials related to this survey: the questionnaire, individual responses, and collated results, are available in WIPOSTAD at: [http://www.wipo.int/wipostad/en/surveys/ipo\\_practice/](http://www.wipo.int/wipostad/en/surveys/ipo_practice/) under the title “OCR practices”.

5. *The CWS is invited:*

*(a) to note the information  
contained in the Annex;*

*(b) to consider and approve the  
publication of the Survey Report in  
WIPOSTAD.*

[Annex follows]

## SURVEY RESULTS

### INTRODUCTION

1. The present survey was conducted in 2011 on the basis of the questionnaire approved by the Standards and Documentation Working Group (SDWG) in October 2009 within the framework of the revision of WIPO Standard ST.22 adopted in November 2008. (See Task No. 37 in the Annex to document CWS/1/9 and paragraph 52 of CWS/1/10 Prov.)
2. The questionnaire addressed issues concerning WIPO Standard ST.22 (Recommendation for the authoring of patent applications for the purpose of facilitating optical character recognition (OCR)) and patent applications submitted on paper or submitted electronically (e-filed) but having the text body of the application submitted in image form (e.g., PDF or TIFF images), as well as questions on OCR practices of Industrial Property Offices (IPOs).
3. The following 30 Offices participated in the Survey:

AT	Austria	IT	Italy
AU	Australia	JP	Japan
BR	Brazil	KR	Republic of Korea
BY	Belarus	KZ	Kazakhstan
CN	China	LT	Lithuania
CR	Costa Rica	MD	Republic of Moldova
CZ	Czech Republic	PL	Poland
DE	Germany	RU	Russian Federation
DK	Denmark	SA	Saudi Arabia
ES	Spain	SE	Sweden
GB	United Kingdom	SK	Slovakia
HR	Croatia	UA	Ukraine
HU	Hungary	US	United States of America
IE	Ireland	WO	World Intellectual Property Organization (WIPO)
IL	Israel		(International Bureau of)
IS	Iceland		

4. The report presents the summary of responses grouped by sections of the questionnaire. Throughout the present document, certain comments have been reworded from the original individual responses for the purpose of abbreviation, clarification, and/or harmonization. Any deviation in meaning from the original comment was not intended.
5. Individual IPO responses are published separately in the original language (the language of the response), along with the automatically collated results, in WIPOSTAD.

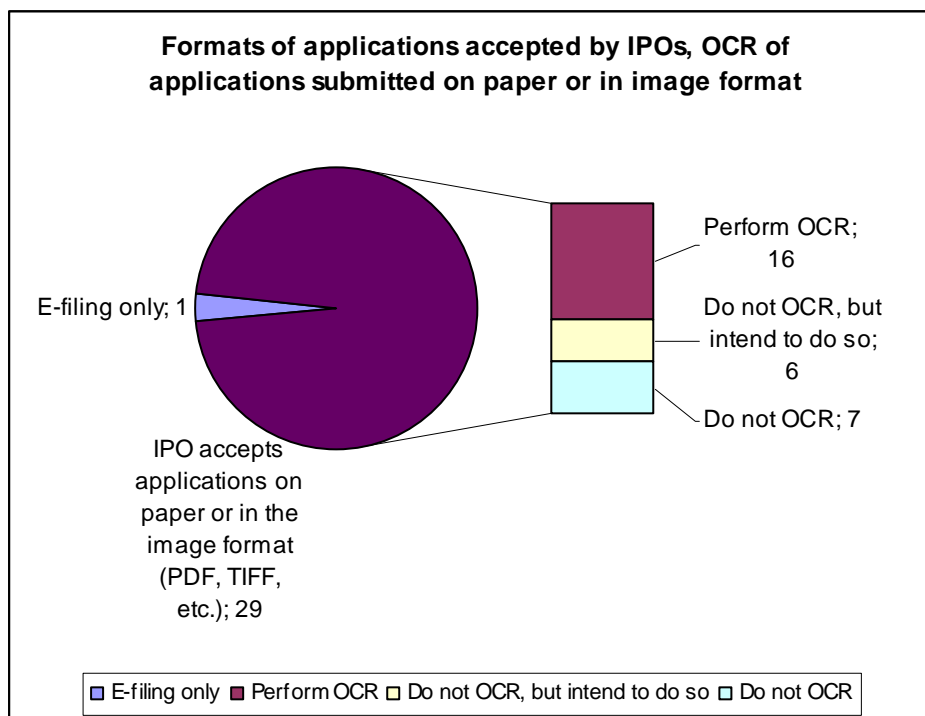
### PATENT FILING

6. The first three questions formed Section 1 of the questionnaire; they focused on the procedure, statistics and format in which patent applications are accepted in IPOs. The answer rate to these questions was 100% (30 offices).

7. The vast majority of IPOs (29 out of 30 offices) reported that they accepted patent applications submitted on paper or submitted electronically but having the text body of the application submitted in image form; 16 of them perform OCR on patent applications and

6 of them intend to do so in the future. One Office (KR) reported that no applications on paper or in the image form were accepted, it was indicated in the response that OCR was performed in this Office though.

8. The graph below shows the distribution of responses on OCR practices in the Offices accepting paper and image format patent applications.



Option	Responses
Perform OCR	AT, AU, BY, CN, CZ, DE, ES, GB, HR, HU, KZ, PL, RU, SE, UA, WO (16)
Do not OCR, but intend to do so	BR, CR, DK, IL, LT, SK (6)
Do not OCR	IE, IS, IT, JP, MD, SA, US (7)
E-filing only	KR (1)

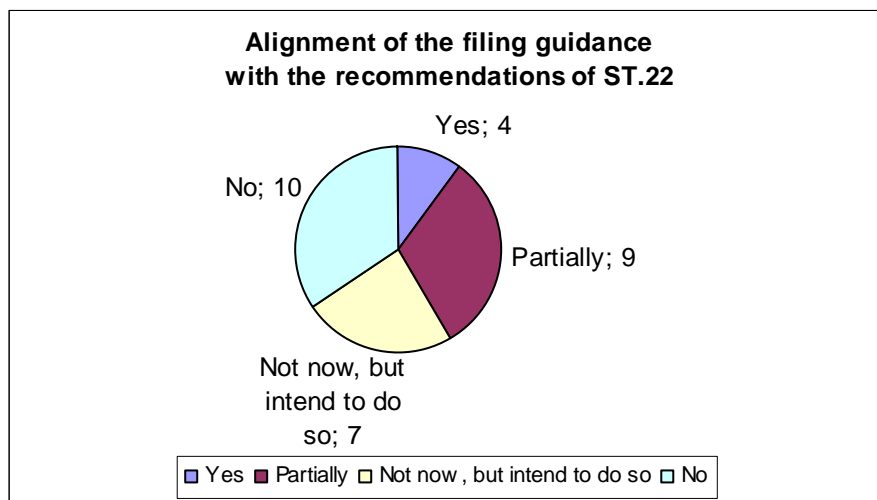
9. The percentage of paper filings, as reported by IPOs, is widely spread and vary from 2.4 % (JP) to 100% (LT, BR, MD, BY, SA and CR), as well as the percentage of “image-format” filings, which vary from 0.01% (UA) to 90% (US, DK). At the same time, it should be noted that for the majority of the responding offices (23 out of 30 offices, 77%) the total percentage of such applications (filed on paper or in image format) is 100%, this means that for these IPOs all received applications may be considered as a possible object for OCR.

10. It was reported by 57% of respondents (17 out of 30 offices) that they performed OCR on patent applications, some others (6 out of 30 offices, 20%) intend to do so in the future, the rest 7 offices (23%) do not perform OCR and have no plans to introduce it.

**PROMOTION AND USE OF WIPO STANDARD ST.22**

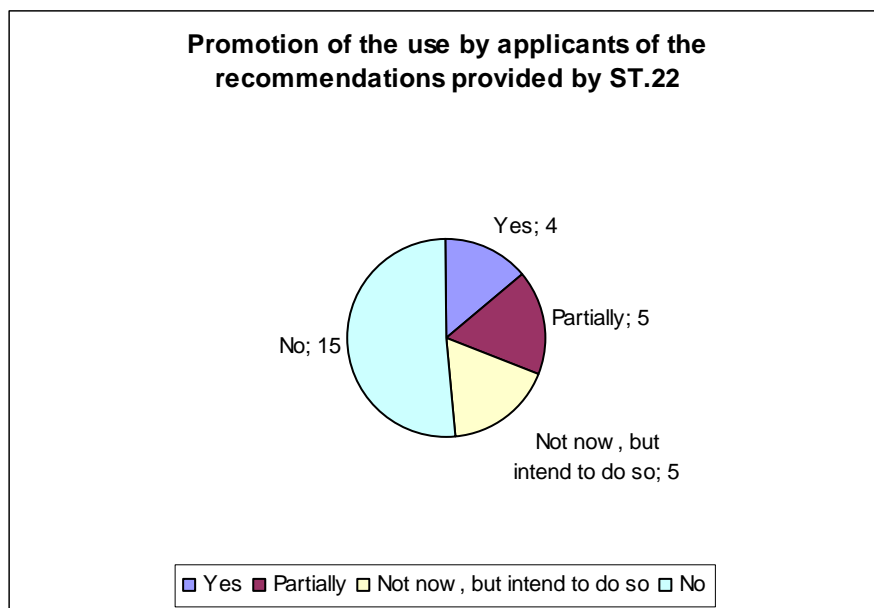
11. Next four questions (Questions 4 - 7) formed a section related to increasing of public awareness on the recommendations of WIPO Standard ST.22 and encouraging applicants to follow them. The answer rate to these questions was 100% (30 offices).

12. Two thirds of IPOs responded that they had adapted filing guidance, which were completely (4 offices) or partially (9 offices) in line with revised recommendations of WIPO Standard ST.22, or that they intended to do so in the future (7 offices). Statistics and distribution of responses are represented on the graph and the table below.



Option	Responses
Yes	AU, BY, GB, WO (4)
Partially	AT, DE, ES, HU, KR, LT, PL, RU, SA (9)
Not now, but intend to do so	BR, HR, IL, IS, IT, SK, UA (7)
No	CN, CR, CZ, DK, IE, JP, KZ, MD, SE, US (10)

13. Fifty percent of IPOs (15 out of 30 offices) communicated that they had promoted the use of ST.22 recommendations between applicants completely (5 offices) or partially (5 offices) or intended to do so in the future (5 offices). Other half (15 out of 30 offices) do not have such plans. Statistics and distribution of responses are represented on the graph and the table below.



Option	Responses
Yes	AT, AU, BY, GB, WO (5)
Partially	ES, KR, LT, PL, RU (5)
Not now, but intend to do so	DK, HR, IL, SK, UA (5)
No	BR, CN, CR, CZ, DE, HU, IE, IS, IT, JP, KZ, MD, SA, SE, US (15)

14. Several IPOs (AT, AU, BY, GB, PL, RU and SK) provided examples of URLs of announcements on the revision of ST.22.

15. Apart from the adaptation of filing guidance or corresponding regulations and their publication, IPOs use the following means of ST.22 promotion: website link to ST.22, translation of the Standard, or parts thereof, into the national language, information circulars, publications in IP Magazine, training courses, advisory services, etc.

#### IMPLEMENTATION OF WIPO STANDARD ST.22

16. Six following questions (Questions 8 - 13) related to the IPO practices of ST.22 implementation. These questions were asked to the respondents that had indicated that they had promoted the use of ST.22 by applicants fully or partially (10 offices, 33%). Six IPOs answered to these questions, namely: AT, BY, GB, ES, RU and WO. The AU Office commented that it was premature to estimate an improvement since the corresponding Guide was implemented one month earlier (in July 2011).

17. The majority of respondents noticed some improvement in the quality of the applications submitted following the recommendations of ST.22 with respect to their formal presentation and layout. Thus, noticeable improvement was indicated by BY, ES and RU Offices. The RU Office sees the reasons of the improvement in the common applicant computerization, rather than in implementation of new Regulations. Little improvement was noticed by AT and GB. The AT Office specified that new Regulations prevented applicants from using too small font size for text parts of the application; the WO reported that no improvement was noticed since it was not measured.

18. It was not possible to draw any relevant conclusion with respect to the changes of OCR quality and costs resulting from applicants' awareness of WIPO Standard ST.22. One respondent (BY) indicated a noticeable improvement of OCR quality, while two respondents (GB and ES) noticed little improvement and half of responding Offices (AT, RU and WO) did not notice any improvement. It was though commented by the RU and WO Offices that no corresponding statistics was collected and the AT Office stated that the applications in question were not OCRed yet. Noticeable decrease in the OCR costs in terms of work-time saving was indicated by the BY Office, ES Office noticed little decrease of OCR costs. Other four Offices (AT, GB, RU and WO) reported that no decrease was noticed in this respect.

19. The practices of requesting replacement sheets on the basis of non-conformity of the application to ST.22 are different: some of the respondents (4 out of 7 offices, 57%) indicated that they requested them, others did not. In accordance with the comments received, the general practice is to request the replacement sheets on the basis of national, or PCT, regulations, wherein the recommendations of ST.22 are incorporated fully or partially.

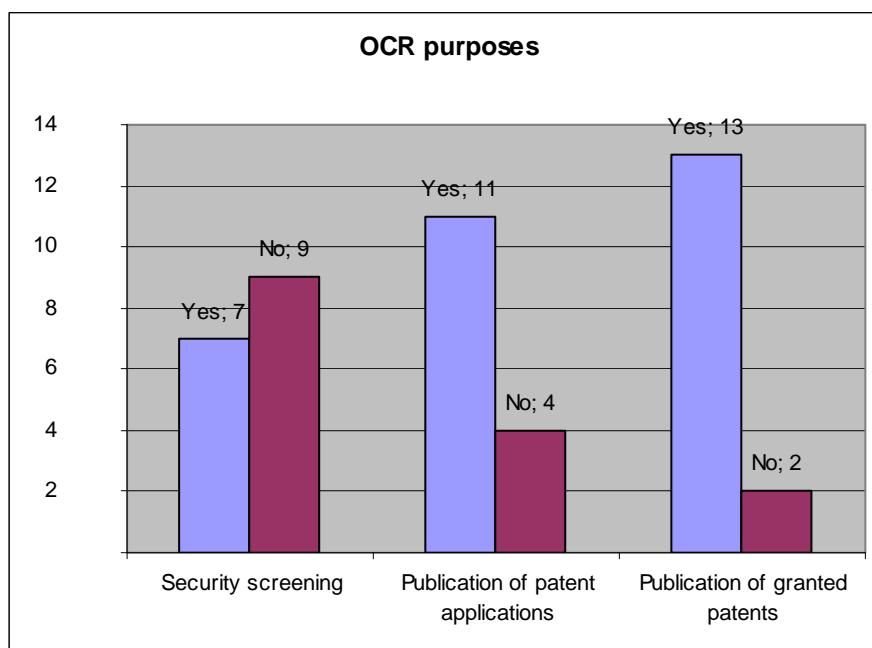
20. Three offices communicated the approximate percentage of applications for which replacing sheets are requested. In the BY Office the replacement sheets were requested for 15% of patent applications, in the GB Office this percentage is 10-15% and in RU Office it was 3-5% during the second half of 2010 and the first half of 2011.

21. No office expressed the intention to take into account the level of compliance of the submitted application with ST.22 for the calculation of fees; GB office commented that this concept seemed interesting though.

#### OCR PRACTICES OF IPOS

22. Following five questions (Questions 14 - 18) related to general aspects of OCR practices implemented in IPOs, like various purposes of OCR, outsource, accuracy requirements and quality checking measures.

23. Sixteen offices answered to the question, or some sub-questions, concerning purposes of OCR, their responses are summarized on the graph and a table below.



Option	Security screening	Publication of patent applications	Publication of granted patents
Yes	BY, CN, DE, HU, KR, RU, UA (7)	AT, AU, CN, GB, HU, HR, KR, RU, SE, UA, WO (11)	AT, AU, BY, CN, CZ, HR, HU, KR, KZ, PL, RU, SE, UA (13)
No	AT, AU, GB, CZ, HR, KZ, PL, SE, WO (9)	BY, CZ, KZ, PL (4)	GB, WO (2)

24. Several offices communicated the accuracy requirements applied for different purposes of OCR. Thus, on the level of security screening of patent application, in the CN Office the accuracy is 99.99%. For publication of patent applications it varies from 99% (AU) to 99.99% (CN). The GB Office capture abstracts with 99.85% of accuracy, 100% of them are manually checked. The WO Office indicated that on this stage they ensured search quality of OCR, i.e., 99.5%, the publication quality in WO Office is 99.95%. The AT and RU Offices have no accuracy checking at this stage. The accuracy requirements for granted patents applied in the AU Office are 99%. In the AT and the PL Offices 100% of documents are manually checked, RU Office have no accuracy requirements fixed.

25. As other purposes of OCR, the respondents indicated the development of search databases (BY, RU) and loading abstract texts into the internal letter writing system to be edited by technical patent examiners before publication and sending to the EPO (GB). In the SA Office OCR is used for internal notes for patent applications.

26. Over 50 % of respondents (16 out of 28 offices) indicated that no quality checking measures were used to control the quality of the OCRed patent documents. At the same time a significant number of respondents (12 out of 28 offices) reported that such measures were in place. These measures include manual check of OCR text against image in electronic dossier or paper originals (ES, GB, PL, RU, SE and UA), MS Word macros for checking and setting of uniform format, orthographic checking, or checking of incorrect recognition of Latin and Cyrillic characters when they are medley presented (RU), vertical and horizontal word checking, text checking and tag checking of randomly selected documents (CN), OCR confidence reported by the software (WO).

27. The procedure of “semi-automated” quality checking applied in CN Office is outlined in the next section “Software and Hardware used to OCR” (see paragraph 32).

28. The following eight Offices (out of 28 which have responded to the question) OCRed patent documents in foreign languages:

IPO	Foreign languages OCRed
DE	No indication of specific languages was provided
HU	English
KR	English
RU	English
SE	Documents regarding European patents validated in Sweden in English, French and German
UA	English, German, French, Spanish, Italian, Greek



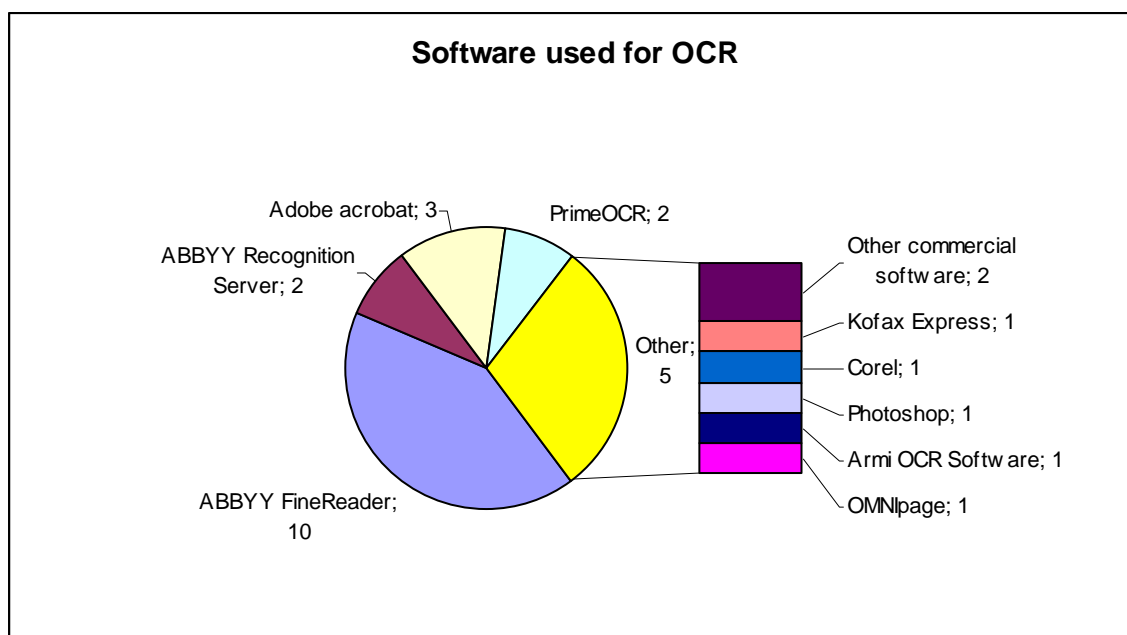
IPO	Foreign languages OCRed
US	Korean, Swedish, German, Chinese, French, Italian, Spanish, Japanese and Portuguese
WO	English, French, German, Spanish, Portuguese, Korean, Chinese, Japanese and Russian.

29. Over 70 % of respondents (20 out of 28 offices, 71%) do not outsource the OCR of patent documents, others do OCR on different stages of the procedure: as soon as the documents are received (CN, JP), pre-grant (CZ, US) and grant stages (US), before publication (ES), etc.

#### SOFTWARE AND HARDWARE USED TO OCR

30. Next three questions (Questions 19 – 21) related to the equipment and software used by IPOs or their contractors for performing OCR of patent documents. The answer rate to these questions was 83% (25 offices).

31. As it is shown on the graph below, ABBYY FineReader, Adobe acrobat, ABBYY Recognition server and PrimeOCR were the most popular software used in the Offices responded to the survey (for further details see Collated results). Four of them indicated that the specific extensions of the software used were developed in order to automate document processing (UA), make the software user-friendly (JP) and perform XML output (WO).



32. The CN Office communicated that the results of OCR using different software were compared and, if any differences were spotted, the document was forwarded to manual check.

33. The GB Office reported that it used the print function in Madras-Phoenix (e-case management system) to capture the Abstract image as a PDF which to then opened and saved using OMNlpage to load into an internal bespoke letter writing system as text for internal technical examiner to amend as appropriate. This text is then used in the publication process.

34. The SA Office does not OCR patent documents since the applicants enclose a soft copy of the application with the application submitted. The Office commented that it planned to switch over to electronic filing in the future.

35. The hardware used to perform OCR of patent documents by IPOs and their contractors is listed in the table below.

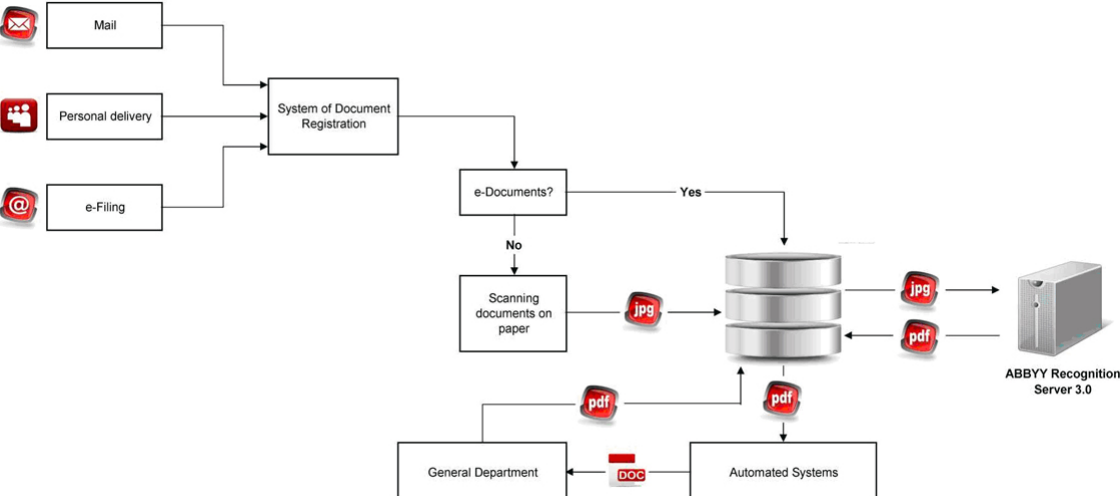
<b>IPO</b>	<b>Hardware</b>
AT	Clients with Windows XP
RU	PC HP 3GHz Documents scanned with Kodak i620 Scanner and Fujitsu fi-5750C hardware go for OCR procedure.
CN	SIPO applies standard PCs to the OCR of patent documents.
UA	Fujitsu image scanners (e.g. 5530C)
US	USPTO hardware
JP	Windows PC
SE	Windows server, Windows Vista
HU	FUJITSU FI-6670/6770A SCANNER
KZ	scanners
HR	PC, scanner
WO	Linux PC servers
AU	Scanner
BY	hp scanjet 5590, hp scanjet automatic document feeder
PL	Scanners:Microtec S400, Fujitsu fi-6230, Fujitsu fi-5900C, Microtec I900
SA	The Office has professional scanners that could be used to perform the OCR of patent documents. They are used for scanning all applications in TIFF format for publication purposes.
KR	HP DL580 G5(P4 Xeon)
ES	Application server

#### WORKFLOW

36. Eight following questions (Questions 22 – 29) related to the workflow of the OCR of patent documents, their storage, subsequent corrections, interrelations of OCR with other components of document processing, use of the documents OCRed by other IPOs and usages of OCRed documents by customers. The answer rate to these questions was 60% (18 offices).

37. Fifteen Offices provided the description of the workflow used for the OCR of patent documents (see the table below).

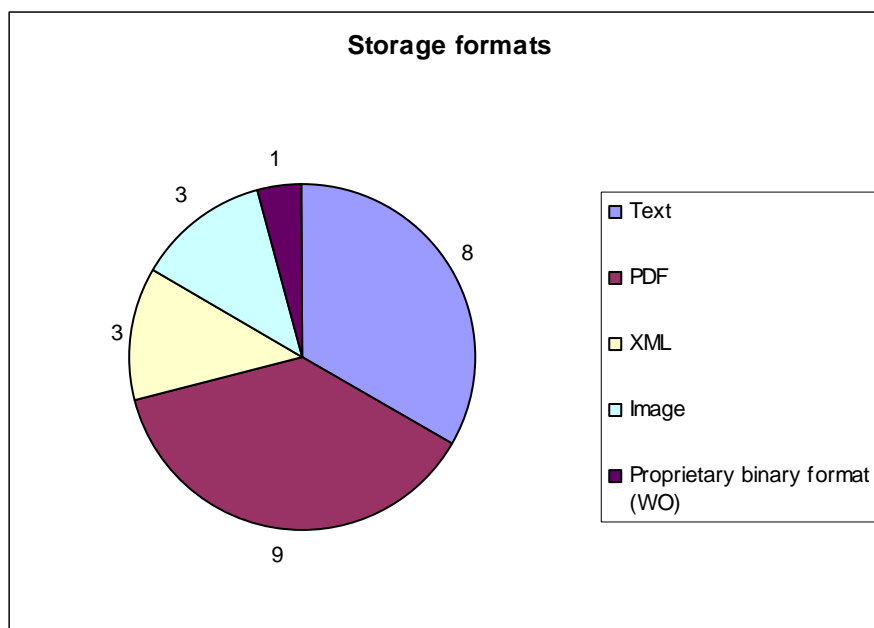
IPO	OCR workflow
AT	<ol style="list-style-type: none"> <li>1. Examiner defines pages for publication.</li> <li>2. Indicated pages are scanned.</li> <li>3. OCR + spelling check + manual formatting in MS Word.</li> <li>4. Formatted text is compared with original paper pages.</li> <li>5. Errors are corrected.</li> <li>6. MS-Word version is transformed to pdf.</li> <li>7. Produced PDF is merged with the PDF of the first page (which has been prepared separately) and (for Utility models) with PDF of the Search report.</li> </ol>
BR	<p>After publication and indexation, all patent documents published since August 1, 2006 are being sent to WIPO's PATENTSCOPE, the agreement foresees the OCR for this documentation, to begin shortly</p>
BY	<p>The claims are scanned after the preliminary examination is finished; bibliographic data are not scanned; other parts of a patent application are scanned for official publication of a patent.</p>
CN	<p>The OCR workflow includes eight procedures: scanning, recognition, vertical work check, horizontal word check, text check, tagging, tag check and quality check</p>
ES	<p>The Office scan documents submitted on paper and send them in electronic format via FTP to the contractor for OCR.</p>
GB	<p>The IPO OCR (captures and converts) abstract text when the applicant requests a Search (within 12 months of filing). This abstract text (when amended by the technical examiner) is used in the publication process (loaded into EPOQUE) if the application proceeds to Publication.</p> <p>Post Publication: the EPO (19-20 months after filing), through a third part agreement, loads GB Full text (Description &amp; Claims) into EPOQUE and Worldwide Esp@cenet databases.</p>
HR	<ol style="list-style-type: none"> <li>1. Scanning documents.</li> <li>2. Creating PDF files.</li> <li>3. Inputting PDF files into OCR software (FineReader).</li> <li>4. Marking all parts of documents (text, tables and images).</li> <li>5. Reading marked blocks.</li> <li>6. Saving doc files.</li> <li>7. Checking in MS Word.</li> </ol>
HU	<p>Scanning the documents, storing the documents, automatic OCR in batch every night.</p>
JP	<p>Filed application documents are converted into image data via scanner and into text data via OCR software</p>
KR	<p>The documents are scanned and OCRed, converted texts are corrected</p>

IPO	OCR workflow
PL	<ol style="list-style-type: none"> <li>1. Scan and OCR of the original document (used software: ABBYY Fine Reader 9.0 Professional Edition).</li> <li>2. Preliminary verification of OCR'd text by the Office's staff.</li> <li>3. Creation of the document (used software: MS Word) based on OCR'd text; bibliographic data; images and claims, using fixed template.</li> <li>4. Saving in DOC format.</li> <li>5. Comparison of the DOC document and paper document by PPO's staff - final revision of the DOC document.</li> <li>6. Transformation of the DOC document into searchable PDF using predefined scripts (used software: Microsoft Word, Adobe Acrobat Professional).</li> <li>7. Publication on PPO's website databases and on the Publication Server.</li> </ol>
RU	<p>OCR is carried out by software ABBYY Fine Reader 9.0.</p> <ol style="list-style-type: none"> <li>1. Before the OCR process, scanned document pages are divided for blocks: text, tables or images.</li> <li>2. OCR-ed text is corrected by operator.</li> <li>3. Pages are saved in MS Word in RTF format.</li> <li>4. Each file is named according to the type of application part – abstract, description, claims.</li> <li>5. Then the text is formatted in MS Word.</li> </ol> <p>Mathematic and physical formulas are put in the text as image objects by formula editor Microsoft Equation. Chemical formulas are put as objects by software ISIS Draw.</p>
SE	<p>New patent abstracts are OCR'd every day, and manually controlled thereafter. Published patent applications and patent documents are OCR'd once a week automatically.</p>
UA	<p>Workflow for OCR of patent documents in UA Office</p>  <pre> graph TD     Mail[Mail] --&gt; SDR[System of Document Registration]     PD[Personal delivery] --&gt; SDR     EF[e-Filing] --&gt; SDR     SDR --&gt; ED[e-Documents?]     ED -- Yes --&gt; DB[(Database)]     ED -- No --&gt; SDP[Scanning documents on paper]     SDP -- jpg --&gt; AS[Automated Systems]     AS -- pdf --&gt; DB     DB -- pdf --&gt; AS     AS -- DOC --&gt; GD[General Department]     </pre>
WO	<ol style="list-style-type: none"> <li>1. Automatic OCR in batch.</li> <li>2. Human proofreading of worse cases identified by the OCR software character recognition confidence levels.</li> <li>3. Export of the OCR in XML and HTML.</li> </ol>

38. 60 % of the IPOs (15 out of 25 offices) responded that they performed the quality check of OCR'd patent documents. The checking is performed manually by the employees of the Office on the whole array (AT, GB, PL, RU and UA) or on selected documents (SE and WO).

39. The majority of the Offices communicated that, when the document was found to be defective after publication, they published a correction or republished the document, 80% (16 out of 20 offices) explicitly stated this in their responses. Six of them (AT, BY, CN, GB, HR and SK) indicated that they proceeded according to recommendations of corresponding WIPO Standards (ST. 50, ST.16 and ST.9). Five Offices (CN, CR, HU, KR and US) reported that the correction was carried out upon request of the applicant. The GB Office commented that they reported the errors in the OCR capturing process to the EPO in order to correct the EPO databases. The CN Office communicated that they sent documents identified as defective back to the contractor for reprocessing, the WO Office indicated that they used the services of an external contractor to improve the OCR of backfile published documents.

40. One of the questions in this section, namely Question 25, related specifically to storage of patent documents after OCR. Twenty-five Offices described their practices, answering to this question or some sub-questions thereof. Twenty-one IPOs provided information on the formats used for storage of OCR'd documents (see table and graph below). The survey shows that the two most popular formats used were PDF and text (mainly MS Word), other formats used were image (mainly TIFF) and XML.



IPO	Formats used		
AT	Text (MS Word)	PDF (bookmarked)	
AU	PDF		
BY	Text (MS Word (rtf))	PDF	Image (TIFF)
CN	XML		
CR	Image		
CZ	Text (MS Word (doc))		
ES	PDF		

IPO	Formats used
GB	Text (MS Word, HTML)
HR	Text (MS Word (doc))
HU	PDF
JP	XML (in conformity with ST.36)
KZ	PDF
LT	PDF
MD	Text (MS Word)
PL	PDF
RU	Text (MS Word (doc, rtf)) Image (TIFF)
SE	Text
UA	PDF
US	XML
WO	WIPO uses proprietary binary format containing all the information coming out of the OCR process

41. More than 50% of respondents (13 out of 22 offices) indicated that the storage format(s) that they used allow for later quality improvements of OCR'd patent documents. For further details see individual responses of AT, GB and JP.

42. With respect to the possibility to quickly identify patent documents with OCR defects, 67% of respondents (14 out of 21 offices) answered that a storage format that they used did not allow for such identification. One third of respondents (BY, GB, HU, KR, KZ, RU and WO) communicated that this possibility existed in their practice (see individual responses provided by the GB and HU for further information). The JP Office commented that the practice implemented in the JPO ensured preventing the generation of defective patent documents.

43. Responding to the question concerning the possibility to view or exchange patent documents in different renditions, 60% of IPOs (12 out of 20 offices) reported that the storage format(s) used allowed for it.

44. Nine of 22 Offices (41%) responded that they retained detailed and complete raw information obtained from the OCR process; other 59 % of IPOs provided negative response to this question. In AT Office the "Fine reader document" was kept for some time in order to make it possible to check spelling or redo OCR (see individual response). The CN Office kept information on the position on the page of Complex Work Unit, such as mathematical and chemical formula.

45. Eighteen Offices responded to the question concerning capturing table contents and mathematical and chemical formula in text format, 5 of them (BY, KR, RU, UA, US) answered positively, 8 Offices (AU, CN, CR, ES, HU, JP, PL, WO) answered negatively and 5 IPOs (AT, BR, HR, SA, SE) did not provide a definite answer. So the results of the survey do not allow defining a general trend with respect to capturing table contents and mathematical and chemical formula in text format, since the IPO practices are different. Moreover, for one Office this question may be answered in different ways depending on certain conditions, such as the complexity of the object to be OCR'd (AT), or the possibility to capture the context in the text format (SE).

46. Half of respondents (12 out of 24 offices) indicated that the OCR of patent documents increased the efficiency of the work of the Office. It was commented that the OCR could facilitate the security screening, preparation of notifications and abstract rewriting carried out by examiners (CN), patent searching (US), quick population of the IPO database with searchable patent documents (PL) and quick access to these documents (IL), examination process (BY, SE), it also assists the translation process (abstracts and reports) and is used to complete a full text publication product (WO).

47. Eleven of 26 Offices responded that they OCR other documents but not only patent documents, for example, documents for protection of different types of industrial property rights (utility models, list of goods and services for trademarks, distinctive signs, etc.), nullity texts for opposition (AT), non-patent literature for internal use (RU, SE), amendment documents and observations from applicants (CN), technology transfer contracts (BR), correspondence (MD).

48. According to the collected responses, the OCRed patent documents are available to and used by IPO examiners as well as by general public to perform patent search. Mainly they are published on the web-sites and publication servers of issuing Offices, electronic products prepared by IPOs, in the databases of PI providers (e.g., EPOQUE and Espacenet were mentioned in the responses), or on CD-ROMs (MIMOSA). Detailed information on Offices' experience is available in the Collated report.

49. Eleven of 26 responded Offices (42 %) use OCRed patent documents provided by other Offices. The most "popular" sources are Patentscope and EPO databases, also mentioned: PCT documents entering national phase (AU), Google patents and DEPATISNet (PL), documents provided by ES Office (CR). The "inverse process" is performed in the WO Office: the IB OCR documents from Mexico, South Africa, Morocco, Israel, Brazil, Panama, Cuba, Spain (very old documents), Dominican Republic, ARIPO and Kenya.

## COMMENTS AND CONCLUSIONS

50. The survey confirms that the majority of documents accepted by the IPOs all over the world are submitted on paper or in image format, and, at the same time, the quality requirements introduced for published patent documents and patent information databases are very high. Consequently, the quality of OCR, as the main step in the transformation of "raw" information received from the applicant to the format in which it will be published by the Office or other patent information provider remains a very important issue and will not lose its importance in the future.

51. WIPO Standard ST.22, which objective is to achieve the lowest possible error rate in the step of automatic reading of the text of patent applications whilst, at the same time, still permitting efficient personal reading of the document, is the tool aimed to assist in the preparation of a patent application in a typewritten form suitable for the subsequent production of an electronic digitized record of the contents of the patent application. The survey shows that, directly or indirectly, the recommendations of ST.22 are used, or planned to be used in the future, by the majority of IPOs' to ensure that the quality of paper applications or applications submitted in image format is sufficient for their subsequent transformation into storage and publication format using OCR techniques.

52. Taking into account the above, it can be concluded that the current WIPO Standard ST.22 serves its purpose and provides necessary guidance on preparation of patent applications suitable for subsequent OCR. Therefore, for the moment, no revision of the said Standard is needed.

53. IPOs are invited to use the results of the present Survey as a source of information on OCR practices implemented in other Offices. It may also be useful to consider different approaches taken by IPOs regarding to storage formats used and with respect to interaction with applicants and external contractors in order to ensure the necessary quality of patent information services.

[End of Annex and of document]