

# WIPO



SCIT/SDWG/7/5

ORIGINAL: English

DATE: May 12, 2006

**WORLD INTELLECTUAL PROPERTY ORGANIZATION**  
GENEVA

**E**

**STANDING COMMITTEE ON INFORMATION TECHNOLOGIES**  
**STANDARDS AND DOCUMENTATION WORKING GROUP**

**Seventh Session**  
**Geneva, May 29 to June 1, 2006**

REQUEST FOR THE REVISION OF WIPO STANDARD ST.22

*Document prepared by the Secretariat*

1. During the discussions at the World Intellectual Property Organization (WIPO) and the European Patent Office (EPO) Bilateral Technical Meeting, held in The Hague in September 2005, it was decided that, agreeing that most of the recommendations set out in WIPO Standard ST.22 (Recommendation for the Presentation of Patent Applications Typed in Optical Character Recognition (OCR) Format) are still valid, it was, however, necessary to revise and update the Standard. This revision would also be an opportunity to reformulate the Standard using the modern information technologies terminology.
2. Consequently, the International Bureau of WIPO and the EPO have prepared a project brief on the revision of WIPO Standard ST.22 for consideration by the Standards and Documentation Working Group (SDWG) of the Standing Committee on Information Technologies (SCIT). This project brief is reproduced as the Annex to this document.

3. *The SDWG is invited:*

(a) *to consider the proposal to revise WIPO Standard ST.22 as requested in the project brief reproduced as the Annex to this document; and*

(b) *to consider establishing a task for the revision of WIPO Standard ST.22 and to set up a Task Force to handle such revision.*

[Annex follows]

ANNEX

PROJECT BRIEF OF THE REVISION OF WIPO STANDARD ST. 22

This proposal is made by WIPO (PCT Information Systems Division) and the EPO (Directorate Document Acquisition).

Indication of the problem or specific need

1. WIPO Standard ST.22 “Recommendation for the presentation of patent applications typed in optical character recognition (OCR) format” was promulgated in December 1990. While most of the recommendations it contains are still valid, some of them need to be refreshed, clarified and validated, and some new ones need to be added, taking into account:

(a) the significant technical advances in the OCR field and the new electronic standards that have emerged (such as UNICODE, XML);

(b) the changes in the way patent applications have been authored, submitted and published during the last 10 years; and

(c) other recommendations on the same subject promulgated separately by other offices for harmonization.

2. This is considered a suitable time for the revision since the major industrial property offices (IPOs) have been, or are in the process of, launching vast OCR projects focused on obtaining the best possible accuracy. Experience shows that if documents follow simple layout rules, the automatic OCR procedures are sufficiently effective to yield a satisfying result for full text search purposes (i.e., an average accuracy rate above 98.5%). However, applications having pages with complex layouts or unusual characters require complex and expensive human corrections to achieve a comparable accuracy rate. In a world where full text publication is becoming standard practice, this recommendation aims at obtaining quality full text searchable published documents for applications submitted on paper or in image-based formats at a reasonable cost for the IP community.

How the need was determined

3. The need was agreed during the EPO-WIPO Technical Bilateral Meeting in The Hague on September 23, 2005.

### Objective of the proposed task

4. The objectives concerning this request are as follows:
  - Redefine the applicability of the Standard in the world of electronic documents; notably, the title of the recommendation would need to be changed to a title such as: “Recommendation for the Authoring of Patent Applications to be submitted in image form or on paper for the purpose of facilitating Optical Character Recognition (OCR)”;
  - Refresh the recommendations to take into account the progress in OCR technology and document electronic representation, and in patent authoring software;
  - Reformulate the recommendations so that the Standard does not appear outdated at first reading;
  - Bring the Standard to the attention of applicants and offices to improve its effectiveness and facilitate offices’ OCR procedures, both as far as quality and costs are concerned.

### Options for solutions

5. The existing recommendations need to be reviewed and a list of additional recommendations need to be discussed and agreed upon. The Standard needs to be reformulated using modern information technology terminology.

### Expected benefits

6. The expected benefits are the following:
  - Reduced costs for OCR procedures performed by offices.
  - Better quality for the full-text published documents built from OCR procedures.
  - Better representation in XML for the full-text published documents built from OCR procedures.
7. The cumulative effect of the above-mentioned benefits is expected to lead to improved search mechanisms for both offices and applicants.

[End of Annex and of document]