

WIPO



SCIT/ITPWG/2/5
ORIGINAL:English
DATE:December17,2002

E

WORLD INTELLECTUAL PROPERTY ORGANIZATION
GENEVA

STANDING COMMITTEE ON INFORMATION TECHNOLOGIES INFORMATION TECHNOLOGY PROJECTS WORKING GROUP

Second Session
Geneva, February 3 to 5, 2003

CLAIMS PROJECT STATUS REPORT

Document prepared by the Secretariat

Background

1. CLAIMS was approved by the Standing Committee on Information Technologies (SCIT) Information Technology Projects Working Group (ITPWG) in its first Session in September 2001. The Classification Automated Information System (CLAIMS) is a project closely related to the progress of the IPC reform, the main goal of which is to enhance the searchability of patent information on the basis of the use of the reformed International Patent Classification (IPC) and its advanced level.
2. CLAIMS will provide links to the global Master Classification Database to be set up by the offices of the Trilateral Cooperation. It will provide tools for computer-assisted classification (pre-classification and re-classification) of patent documents. Translation memories should be developed to support the translation of the advanced level of IPC into French, and to support the translation into other official, but not authentic languages. These tools should have an interface to the current IBIS system ("virtual IPC"). Interactive IPC tutorials will help better understanding of the reformed IPC.
3. The previous International Patent Classification (IPC)-related project, IBIS, is the foundation for the CLAIMS activities (see documents SCIT/ITPWG/1/13 and PBC/3/3).

4. The present document contains a progress report on the development of the CLAIMS Project.

Overall Project Status

5. The project has experienced some delays due to certain reasons. However, it is now running well, and started producing results that go beyond the original expectations.

Summary of Progress

6. Major accomplishments include:

(a) The selection of three companies that provide resources for the coding and data processing for the project;

(b) In the area of Automatic Categorization, the final scope of the project has been defined; two categorization programs have been tested with a first set of training and test data created ad hoc, with promising results;

(c) In the area of translations support, several translation memory tests were carried out;

(d) As for IPC tutorials, the work will be finished in January; training examples have been created and stored in a relational database; and

(e) With respect to the IPC reform support, further progress has been made toward the conversion of Catchword Indexes in various languages into XML language and toward the integration of chemical illustrating formulae into one single source. The Data Entry Interface for the Definitions has been elaborated.

Project Milestones/Deliverables Planned and Achieved (for the past three months)

7. As for the coding and data processing, the Methodology to be used has been established and used: it is based upon the Rapid Application Development methodology, adapted to WIPO's environment. Programming started on October 1, 2002.

8. In the area of automatic categorization, a survey on the state of the art of the automatic categorization of patent documents was completed and published on the web for the International Patent Classification (IPC) reform community; the scope of the project was defined together with the sponsors; an alpha set of the IPC training and test using the PCTi2C data (the full -text archive of the PCT applications) was produced; two categorization programs were retested and re-run on the EPO pre-classification data with promising results.

9. In the area of translations support, the usefulness of the translation memories through several memory translation tests has been demonstrated.

10. As for IPC tutorials, several IPC training examples were created and stored in a relational database that allows their extraction for use online and their export for off-line use.

11. In the area of IPC reforms support, the XML converted Catchword Index was integrated into IBIS; the German unofficial Catchword Index was converted into XML; the chemical illustrating formulae were consolidated into a single source; the Data Entry Interface for the Definitions was elaborated.

Project Milestones/Deliverables for the Next Period (three months)

12. The following deliverables are foreseen for the next three months:

- (a) The IPC tutorial package will go into production by February;
- (b) The data originated by the second revision, which was received only recently, will be integrated into the system by January; and
- (c) The external tests on the alpha IPC (training and test) set are expected to be complete by the end of February 2003 and the creation of a beta set and delivery of test results is expected by the end of March 2003.

Key Issues

13. There are no special issues to be mentioned.

14. The ITPWG is invited to comment upon and note the contents of this document.

[End of document]