

WIPO



PCT/TCO/SS/III/14
ENGLISH ONLY
DATE: July 3, 1972

WORLD INTELLECTUAL PROPERTY ORGANIZATION UNITED INTERNATIONAL BUREAUX FOR THE PROTECTION OF INTELLECTUAL PROPERTY GENEVA

PATENT COOPERATION TREATY

INTERIM COMMITTEE FOR TECHNICAL COOPERATION

STANDING SUBCOMMITTEE

Third Session, Geneva, October 2 to 5, 1972

PROGRESS REPORT ON INIS

Report by the International Bureau

SUMMARY

This document constitutes a Progress Report on the International Nuclear Information System, particularly in respect of its suitability for use by Patent Offices for searches in non-patent literature in the nuclear field.

Introduction

1. A meeting was held in Geneva on November 5 and 6, 1971, of the Joint ad hoc Group for Patent Information in Nuclear Science and Technology. Represented at this meeting were the International Atomic Energy Agency (IAEA), national organizations for nuclear energy, WIPO, national Patent Offices, and the International Patent Institute (IIB). During this meeting it was agreed that to assist Patent Offices in making full use of the International Nuclear Information System (INIS) of the IAEA as a search tool particularly for non-patent literature the IAEA would invite WIPO to send representatives, together with representatives of Patent Offices, to meetings arranged to familiarize potential users with INIS (see document PN/I/4, paragraph 13(b)).

2. At the invitation of the IAEA, WIPO and the IIB each sent a representative to an INIS Indexing Training Course in Vienna from June 19 to 23, 1972. The present document is a report on INIS by the representative of WIPO.

The System

3. INIS uses a coordinate indexing system which employs a Thesaurus of terms based on preferential, hierarchical and affinitive relationships. The latest edition of the Thesaurus (IAEA Publication IAEA-INIS-13(Rev.2)) contains nearly 11,500 descriptors (accepted terms) and over 3,000 forbidden terms. The terms are alphabetically arranged and the descriptors may each be followed by scope notes, and lists of related forbidden terms, broader terms, narrower terms, and related descriptors.

4. Records of newly published literature in the nuclear field are prepared for input at national and regional centers (at present 32 countries and two international organizations provide input to the system) and are sent, in the

form of worksheets, paper tape, or magnetic tape, to the IAEA where they are processed by computer and merged to create a master file. The computer input is automatically checked for compliance with the descriptive cataloguing rules, for correctness of the subject categories, and for the existence in the Thesaurus of the descriptors assigned. Minor errors, such as a single spelling error in a descriptor, are automatically corrected, and other errors are flagged for manual correction. The computer also flags the use of fewer than six or more than 19 descriptors for an item. In addition, an "up-posting" procedure is carried out, whereby for every descriptor assigned all of its associated broader terms are automatically assigned.

5. Magnetic tapes containing the bibliographic data and descriptors for each item processed during each month are distributed to Liaison Officers in 52 national centers and international organizations subscribing to the service. In addition, the same information is available in printed form in the monthly publication "INIS Atomindex" at an annual subscription fee of \$25. Items are numbered in a continuous sequence which started at 00001 in May 1970. Each issue contains a personal author index, a corporate entry index and a report and patent number index. Cumulative indexes are issued twice yearly.

6. For each item entered into the system an abstract is provided. In the majority of cases (at a rough estimate about 90%) an author's abstract, or a translation of it, is used. The abstract must be in one or more of the official languages of the IAEA (English, French, Russian and Spanish) and in fact they are almost invariably in English, sometimes supplemented by additional versions in one or more other languages. The abstracts are available only in microfiche form and are identified and arranged according to their sequence numbers in "Atomindex". The subscription rate is \$45 per annum.

7. Approximately 30% of all items entered into the system are from so-called non-conventional literature, i.e., literature which is not widely distributed or is not readily available. Included in this category are technical reports, conference preprints, theses and patents (although patent gazettes are regarded as conventional literature). Copies of the full texts of all items of non-conventional literature reported in the system are available from the IAEA in microfiche form. Microfiche of individual items may be ordered at \$0.65 per item. Alternatively a complete supply can be obtained on a standing order at \$0.20 per microfiche.

8. It is estimated that there are some 80,000 literature items per year dealing with the peaceful uses of nuclear science and technology. Initially INIS restricted its scope to a small number of the more important areas, but early in 1972 was expanded so that it now accepts input from the whole field on a voluntary basis. From January 1973 all inputting centers will be expected to cover the whole of the nuclear field on an obligatory basis. At present the number of reported items per month is approximately 700. Patent and like documents will be included, but to date no more than a handful have appeared in the system.

The Indexing Training Course

9. Apart from the opening and closing sessions, the Training Course was divided into four parallel courses directed towards Physics, Chemistry, Engineering and Life Sciences. In each of these courses participants were given a copy of a document relevant to the course subject and were allowed 20 minutes in which to read the document, select an appropriate subject category (or categories), and index it using the Thesaurus. At the end of the time limit, participants were invited to reveal the descriptors which they had assigned. A discussion then followed on the merits and demerits of each descriptor assigned and eventually an "optimum" list was drawn up. At some stage of the discussion a list of descriptors assigned by the lecturer (an experienced INIS indexer) was introduced.

10. It was clear that there was considerable variation between participants regarding the descriptors selected. It was impossible to obtain an accurate estimate of the indexing consistency, but on average the indexers probably assigned correctly perhaps ten out of fourteen descriptors in the optimum list. The lecturers usually revised their own lists as a result of the discussions, two or three changes being typical. It was noticeable that many of the participants were not always entirely convinced by the arguments put forward by the lecturer for or against the use of a particular descriptor.

11. Towards the end of the course participants were given advice on retrieval strategy. The IAEA uses the IBM Information Retrieval Management System and suggests that other users might do likewise. The well-known relationship between

recall and relevance was explained to participants but matters such as profiling, screening of the output, and usage of the INIS microfiche files were not discussed. It was not possible to evaluate the effectiveness of INIS for search purposes using the descriptors.

The Abstract

12. In order to obtain some idea of the value of INIS abstracts for patent searches, a small informal test was carried out. From the list of subject categories (approximately 130, including main series), 22 categories were selected as being most likely to contain patent-relevant material. The latest available issues of "Atomindex" (April and May 1972) were then examined and from each of the 22 categories selected a further selection was made of those items within each category which from the title and the list of descriptors appeared most likely to be patent relevant. Finally the corresponding microfiche were scanned and the abstracts of the selected items were read. The selection was made mostly from the May issue of "Atomindex", the April issue being used only to find examples in categories not represented in the May issue. Probably the effective number of items in the sample was of the order of 800, of which 57 were selected as being most likely to be patent relevant. Examination of the microfiche showed that 13 of the abstracts were sufficiently informative to be of possible value in a patent search. The remainder were very general, disclosing no more than the nature of the subject under investigation, were theoretical in nature, or gave only experimental results.

13. The two issues of "Atomindex" which were examined contained a reference to only one patent document. The abstract of this document was one of the final group of 13 considered to be of value in a patent search.

14. It was not possible to examine the full texts of the 57 items selected in order to determine if the items themselves were patent relevant. Such an investigation could usefully be carried out as part of any further evaluation of INIS. This would, of course, require access to a collection of journals on nuclear science.

Use of INIS by Patent Offices

15. While the Indexing Training Course did not permit any quantitative evaluation of indexer accuracy or consistency, or of search results, the general impression gained was that the results of any such evaluation would not be very encouraging. When patent documents are entered into the system on a regular basis a search using INIS descriptors would probably not give as good results as a search carried out using any of the classification systems presently used by Patent Offices in the nuclear field. For non-patent literature INIS will certainly give a very complete coverage of the field when the system is fully operational, but a rather high proportion of the literature covered, and of the literature retrieved in a search, would be irrelevant to Patent Office needs, being concerned with theoretical and experimental aspects without practical application.

16. In order to use INIS successfully it would be necessary for Patent Offices to have access to all the journals from which items are selected for input to the system. The INIS Authority List for Journal Titles (IAEA Publication IAEA-INIS-11 (Rev.0)) contains approximately 2,500 titles, but items may be obtained also from other journals not in the Authority List. Copies of non-conventional literature items would be obtainable, in microfiche form only, from the IAEA.

17. It appears that the INIS abstracts would be of little value in determining whether an item would be of interest in a patent search. They would serve only as a rather coarse screen for eliminating false drops. In the majority of searches it would probably not be worthwhile looking at the abstracts at all--screening could be done as quickly, and more successfully, by looking at the items themselves, particularly since these would be accompanied in almost 90% of the cases by the author's abstract used by INIS.

Conclusions

18. As the Indexing Training Course was directed, as its name implies, to the training of indexers, it was not ideally suited to the evaluation of INIS as a search tool for Patent Offices. It seems, however, that even from this limited

study INIS cannot be recommended at the present time for Patent Office use. Nevertheless, it would perhaps be useful for further studies to be made now, to provide a basis for comparison with later studies and enable future improvements in the system to be evaluated. Such further studies could extend to the investigation of a sample of full texts of items entered into the system, and possibly, with the cooperation of the IAEA, the carrying out of a search test. The results of such a search test could be compared with those of the search test carried out in 1967 by Mr. Dekker of the Netherlands Patent Office on behalf of ICIREPAT (see document IC/TC.I/27(70)). The Euratom system used for that test was a predecessor of the present INIS system.

19. The Standing Subcommittee is requested to take note of the information contained in this document and to express an opinion on the conclusions reached.

/End of document/