INFORMATION COLLECTION STRATEGIES TO SUPPORT STRENGTHENED SAFEGUARDS

XA0200141

L. COSTANTINI
International Atomic Energy Agency, Vienna, Austria

J. HILL
Australian Safeguards and Non-proliferation Office, Canberra, Australia

The IAEA Board of Governors approved the implementation of Part 1 of Strengthened Safeguards in June 1995. Since then, the collection and analysis of information beyond that provided by States parties and acquired by inspectors under NPT Safeguards Agreements has been an integral part of IAEA safeguards. The Agency has formally established internal structures and procedures to facilitate the effective use of open-source and other information not previously used in safeguards.

Over this period the IAEA Division of Safeguards Information Technology (SGIT) has been building its collections of electronically held open source information. Some of these collections are quite nuclear-specific, such as material from the Monterey Institute in California, and nuclear news collections provided voluntarily by a number of Member States. Others are completely general news sources. Several of these collections contain many more reports than could possibly be reviewed, or even skimmed through by a human analyst.

So a need has arisen for computerised search facilities to identify nuclear-relevant items from those collections. More specifically, the need was for search mechanisms to identify reports that would be useful to inspectors responsible for preparing State Files and State Evaluations, and for making the comparisons with declarations needed to identify questions and inconsistencies.

The Agency has more than one piece of software available to it to help with the searching and analysis of substantial collections of reports. Search 97 from Verity was chosen for this particular application because it is very straightforward to use, and it was expected that personnel from all over the Department of Safeguards would carry out these searches on a routine basis. The approach whereby special-purpose search mechanisms are designed for use by a large number of users, who are unfamiliar with the details of the search software, seems to be unusual if not unique to the Agency.

Search 97 employs a function called Concept retrieval, a technology which enables users to search for subjects or concepts in documents, rather than individual words or phrases. Search 97 treats specific words and phrases as evidence for the existence of a concept. Search 97 encapsulates its search terms in a component called "TOPIC tree". In designing Topic Trees, each concept is subdivided in different sub-concepts and be weighted because some words, phrases or other concepts are more important than others in expressing a concept.

The task began with the use of Search 97 to find reports dealing with each of the proliferation-relevant fuel-cycle processes. The Agency's Physical Model describes each of the nuclear fuel-cycle processes that a state with a nuclear weapons program might need, from the mining and milling of uranium onwards. At each stage the Model identifies "indicators", which are potential observables that would suggest that a state was undertaking (or envisaged undertaking) the fuel cycle step in question.

It is particularly appealing to use Search 97 with the Physical Model, because both require the presentation of information in a tree-like form, proceeding from the most general information (for example indicators, or key words, relevant to all uranium enrichment processes) at the top level of the tree (on the trunk). At the second level of generality (or on the biggest branches of the tree, coming from the trunk) we would find indicators or key words relevant to major groups of processes (for example enrichment using UF_6, enrichment using UCL_4 and enrichment using uranium metal). And at the lowest level of the tree we find indicators or key words specific to individual processes such as centrifuge enrichment.

From this point on, derivation of a TOPIC tree diverges from the Physical Model, because a key word is very different animal from an indicator. Synonyms for key words need to be introduced, together with conditions governing the return of reports if a key word is not entirely specific to the process under consideration. Much of this work is done while testing the tree against real collections. So the initial "top-down" approach becomes more "bottom-up" in nature. Nonetheless, the Physical Model provides a very useful starting point.

A series of TOPIC trees has been designed and introduced into routine service in the Department of Safeguards. As well as the processes covered by the Physical Model, these trees cover certain subject headings in the standard State File structure, and other subjects of interest to the Department. Quality Assurance procedures have been introduced whereby the performance of the trees is checked against a "representative" collection of reports, amongst which the reports dealing with the subjects covered by the trees have been identified in advance. It is also possible to assess their performance by comparison with search procedures used by other organisations.

SGIT's mandate is to collect, store, process and disseminate open-source and non-safeguards information for use in the overall country-by-country evaluation process and for use in other products, such as ad-hoc reports written at the request of various high level Agency officials. SGIT uses the above mention TOPIC trees to organize, cull, and finally select relevant open-source information for State Evaluation Reports.

The expectation is that the State Evaluation Group, comprised of country officers in Operations Divisions and its support Divisions, will treat the reports returned by the TOPIC trees as a rough "first cut" from the totality of the open source material available to the Agency. They would then manually select and use the information they consider being useful for inclusion their State Evaluation Report.

One of Verity's newest tools, "Knowledge Organizer", uses TOPIC trees to categorize and store vast bodies of information into directories with like documents. This makes the process of collecting and disseminating information both more timely and efficient. The Knowledge Organizer also allows SGIT to extract reports dealing with a particular TOPIC tree subject, for a particular country, originating within a particular timeframe. Country Officers typically ask for the most recent reports dealing with countries for which they are responsible. Knowledge Organiser allows such a service to be provided with minimal effort. It also allows SGIT to implement a long sought after product, the electronic state file. SGIT was able to customize the basic Knowledge Organizer feature to develop an electronic state file system. The electronic state file is essentially an encapsulated set of selected files on a particular country, categorized and stored into sub-directories based on TOPIC trees that match differing parts of the of the nuclear fuel cycle. The electronic state file interface allows the State Evaluation Group to review, search, and select information that will be useful for conducting the state evaluation report.

ACKNOWLEDGEMENT