

A NEW DIMENSION TO RETRIEVAL

Mike J. Collins
INSPEC, Station House, Nightingale Road, Hitchin,
Hertfordshire, England SG5 1RJ

Abstract

Searching in online databases for specific numerical data and for chemicals by formulae has always been difficult or, at best, complex. Two new additional indexing services were introduced to the INSPEC database at the beginning of 1987 to overcome these problems: Chemical Indexing - for searching all kinds of inorganic compounds and material systems, and

Numerical Data Indexing - for limiting searches to a specific value or range of values of a particular physical quantity. The literature of SN 1987A, the first supernova to be discovered in 1987, is used to demonstrate the powers of these new indexing tools and how they compliment traditional retrieval aids such as classification, descriptors, and identifiers.

1. Introduction

Visitors to the INSPEC stand at exhibitions and conferences round the world often ask what is "new" on INSPEC in addition to the 240,000 records we at present add to the database each year. Usually we have news of some new product derived from the latest section of the file and a few leaflets to read. Now we have something quite different, something extra that is being added to the database that will give retrieval from the INSPEC Database a new dimension: Chemical and Numerical Data Indexing.

2. Background

Previous studies had indicated that although INSPEC is primarily a bibliographic database there was nevertheless a considerable amount of chemical formulae and numerical data on each record that was at best difficult and at worst impossible to retrieve. For example, on some systems it was impossible to differentiate between cobalt (Co) and carbon monoxide (CO) because there was no lower case facility. Things were even worse when dealing with semiconductors and metallic glasses where for example one had to consider all possible rotations of AlGaAs and the various notations for mixtures. In the astronomy section we had even experimented with numerical data written into the identifier field in a controlled format but searching in ranges still proved difficult.

3. New Developments

Chemical Indexing and Numerical Data Indexing (NDI) was introduced at the beginning of 1987 by creating two new fields which would appear on appropriate records. All the major online systems that make the INSPEC Database available offer a form of search facility for the new indexing. Some only allow searching of the NDI entries as if they are text strings, but Dialog, ESA/IRS and STN treat the data as numbers permitting retrieval of both spot values and ranges.

4. Applications

The new indexing is primarily intended to act as a fine tuning mechanism during an online search. Consider a literature search on SN 1987A in the Large Magellanic Cloud, the first supernova to be covered by our new indexing. By mid-1988 there were already over 300 papers discussing this object from which to choose. In the past this set could have been reduced to say just the ultraviolet observations by using either classification code (A9580M) or descriptor (ULTRAVIOLET ASTRONOMICAL OBSERVATIONS). Now you can be more specific. Using the NDI facility you can restrict the search to just those papers describing observations around say 300 nm.

However the new indexing could prove to be a powerful retrieval tool in its own right. Consider a subsequent search for any papers discussing the ultraviolet spectrum of atomic oxygen, not just in SN 1987A. This would involve other astronomical, atmospheric, and laboratory spectra. The new Chemical Indexing would allow you to select oxygen as an element and the numerical indexing would again restrict the search to the UV or to a specific wavelength. This could be more efficacious than using the descriptors OXYGEN and ATOMIC SPECTRA in combination with classification codes.

5. Future

There is a small thesaurus containing the 43 NDI quantities and their units. It is possible that this list will evolve. At present for example there are no means of retrieving magnetic field strengths and dates would be a valuable addition although a Julian Day Number would probably have to be used as in variable star work. We may even have to reconsider the permitted range of values in the NDI which, for all practical purposes, is 1.0E-70 to 9.9E+70 (DIALOG can in fact handle values in the range 5.4E-79 to 7.2E+75). A recent paper in Astronomy and Astrophysics discussed the rotation period of the Universe which

is suggested to be longer than $10^{135-155}$ years - you can imagine what the computer said after it had converted this to seconds!