# all the Red-Legged Partridges

*By K. Way*

## . . . or . . .

## 99.5% Retrieval of Scientific Information

The word "retrieve" means "to call to mind again, to find again" or, in hunting terms (obsolete), "to discover again game once sprung, to flush partridges a second time". Perhaps the ancient usage with respect to birds is the one which gives the clearest idea of the modern meaning to documentalists. "Retrieval" has come to signify the flushing again of such beautiful partridges as journals, books, reports which have found obscure resting places in the dense forests of modern libraries. But a new concept has been added, namely that in each operation the partridges found again should be of a specific kind. "99.5 percent retrieval" now means that if there are 1000 red-legged partridges in the forest, 995 of these can be sprung again at will.

The concept of 99.5 percent retrieval of any species is not often met in writings of documentalists. They discuss "relevance" in relation to retrieval[1] and find that to achieve nearly 100 percent retrieval on any topic, you have to be satisfied with practically zero relevance; that is, to get all the red-legged partridges you have to take every bird in the forest. For even 25 percent relevance (only one red-legged partridge among four birds brought in), one must be satisfied with 80 percent retrieval[1] (that is you will still miss 20 percent of the red-legged birds). Actually 80 percent is a good round number for the retrieval performance of conventional indexing systems according to the results of the Cranfield project.[1] But this is not nearly good enough for the modern scientist-compiler. In many cases he *has* to have the documentalists' impossibility—about 100 percent retrieval *and* about 100 percent relevance.

To focus attention on a rather simple, specific situation: On the average there are probably five rather carefully measured values of the half-life of each of the 1200 known radioactive nuclei. Thus altogether there are some 6000 experimental half-life values which have to be collected in order to make a comprehensive compilation of half-lives. Such a compilation is the necessary basis for a list of "best" half-life values which is needed by all users of radioactive materials, nuclear scientists, reactor technologists, health physicists, biologists, etc.

Suppose that, in making the half-life compilation, 1200 measurements, or 20 percent of the total, were overlooked. If such were the case, there would be vehement demands for a more thorough job. One value in five may not seem large, but that one value might be the most precise of all. Even if all the values were of comparable accuracy, who would not want to average five, rather than four, results if it were possible to do so?

Omission of 20 percent of the data from any scientific compilation would certainly not be tolerated by scientists. Would the oversight of 60 values out of 6000 be condoned? It is hard to say. Our experience in the Nuclear Data Project has led us to believe that failure to find 30 values out of 6000, or 0.5 percent, is just tolerable. Even that means some ten polite, but rather cool, letters and perhaps ten personal admonishments during meetings. The authors responsible for the other ten overlooked measurements presumably did not see the compilation. The first twenty probably view the compilation, and all other work of the compilers, with some skepticism. Of course, the next edition of the half-life table could be improved by adding the values known to be overlooked in the first issue; but the compilers would have another completeness problem with the measurements published in the interval between editions.

The fact is that it is only sensible to demand at least 99.5 percent completeness in any scientific data collection. Missing bits of information may make all the difference. The experimental work required to determine these data was arduous and costly. If they are not included, many experiments will be repeated unnecessarily.

The situation is quite different with information desired for generally informative purposes. If you want to find out something about Egyptian scarabs because your Peace Corps niece sent you one for Christmas, you go to the library and cheerfully search Egyptology, beetles, gems, etc.,

Katherine Way is director of the Nuclear Data Project at Oak Ridge National Laboratory in Tennessee.

if you draw a blank on scarab itself. If you could press a button and get references to every work which contained any mention of scarabs, you would not think of doing it. The pleasant tome found by your mildly ingenious approach will tell you just about as much as you really want to know about scarabs.

Many scientists have this easily-satisfied attitude toward a number of topics. They suddenly want to know something about hyperfragments or neutron polarization, but not *everything*, for goodness sakes. They want to be oriented, not deluged.

Actually scientific information exists on four rather easily distinguished levels of organization: (1) new experimental results reported in articles, tables, symposium proceedings, etc.; (2) data compilations; (3) review articles; (4) critical tables, books, or extended treatises. The scientist usually knows on which of these levels he wants to enter the information store in a given field. If a review paper on deformed nuclei is what he is after, one or two references will satisfy him. Fifty research papers on the topic would send him scurrying from the library. Perhaps he gets intrigued by deformations and then wants to compare theory and experiment. At this point he will call for the very latest data compilation. If it turns out that this is a little outmoded, he will begin to demand the latest research papers on the topic. Here he will want everything in the field since a certain date. His needs for completeness will then be the same as the compilers' in order to bring the compilation up to date for his own purposes.

Thus, the percentage retrieval desired by a scientist depends on the level of organization at which he needs information. Apparently little attention is being paid to this point by students and devisers of index systems. Almost none at all is being paid to the high retrieval needs of compilers.

One technique open to them is to start with one of the 80-percent indexes and then to enlarge the coverage by looking up all the citations in the papers found there. This is a time-honored approach. Actually it is so inefficient for scientific results that most compilers prefer to start from scratch—that is, to search tables of contents of journals for titles that sound pertinent, then to scan the papers themselves to see if their topic is treated. Usually it takes a pretty experienced scientist to do this. Of course, measurements of such a straightforward quality as a half-life could probably be located, even if hidden in a footnote, by a non-PhD. But it is not always so easy to index a half-life under the proper nucleus. For example, there are three radioactive Tb isotopes with half-lives of approximately 19 hours. To decide which one was produced in a given experiment often requires a knowledge of thresholds, cross sections, and chemistry. The indexing of nuclear reactions for energy-level information is even more difficult. About which nucleus does a particular experiment give information, the target, the compound, or the product nucleus? Since only someone well-versed in nuclear physics can determine this, it is a waste of time for an untrained person to scan papers for this type of information.

The searching-scanning-indexing process precedes the real scientific work of analyzing, comparing, evaluating. It requires some scientific judgment, but not much. An experienced scientist will do it for a while, but not for long. The information centers that are springing up as a result of the Weinberg report are beginning to find this out. Those in business for a while are acutely aware of it.

The problem is how to achieve 99.5 percent retrieval of scientific information without assigning a number of highly trained scientists to the (for them) unrewarding job of literature searching and indexing. Perhaps a few scientists would do the work for everybody for a year or two if they could then be sure of other opportunities later on. This possibility has even gotten to the "proposal" stage but, at the last minute, no laboratory or institution has been willing to contribute any of the needed highly trained scientists even for a short time.

The only alternative we can see at present is to get authors to attach index or key words to their own papers. They are presumably experienced scientists thoroughly familiar with the contents of their papers. Each one, if guided by a list, would need only a few moments to select suitable labels. Since they want their papers to be in compilations, there would be a drive to pick correct labels in order to ensure the notice of the proper compilers. Achievement of 99.5 percent retrieval for the topics on the list would then be possible. For topics not on the list the situation would be just what it is now. Lists can be changed and adjusted to meet new needs. Perhaps one of the major functions of learned societies should be the preparation and revision of such lists.

Under the system just sketched, every paper that reports a measurement of the half-life of

$^{165}$Tm would be so labeled since half-life and all names of nuclei would surely be on the list.

This proposal has been called a "key-word" plan. Actually there are several other key-word projects in the scientific field but none of them have achieved 99.5 percent retrieval. Perhaps it should be labeled the "99.5 percent-retrieval key-word plan" in order to emphasize its goal and present uniqueness.

The words on the list from which the author makes his selection could be grouped so that he need study only a small area of the list. They could, in many cases, be symbols so that translation problems would be minimal. Of course, they should be translatable into machine language, but this seems to present no problems.

Just how such a scheme might work is shown by the example (Fig. 1) taken from *Nuclear Physics*, a journal which is already experimenting with the idea.

One of the main points to be noticed is that the key words bring out points not mentioned in the abstract, indeed not really suitable for an abstract, which should stress the main focus of the paper and its implications for development of the field. Without the key words one would have to read the paper to discover that it contains the most accurate half-life values obtained to date for $^{57}$Ni, $^{62}$Zn, $^{66}$Ga, and $^{67}$Ga. The words given above were chosen from a list, part of which is shown in Fig. 2.

The function of the key words is to index the paper for those topics on which compilations are being made or should be started. If the main purpose is to help compilers, one may ask why add them to the paper? Perhaps they could just be sent by editors to compilers. This is certainly a possibility, but it also seems they would be a great help to the conventional indexers. Moreover, new compilers are always springing up. Indeed, they should be if all the scientific information which is pouring out of the labs is to be analyzed and digested. One-shot compilers often make the most valuable contributions of all. Then there is the compiler who never tries to get into print but just decides some afternoon that he wants to collect all the information on angular distributions of inelastically scattered alpha particles. For all such people, the key words will be a blessing. To many others they may well remain a mystery. As Abelson points out, "For those engaged in efforts at the leading edges of scientific discovery, information retrieval presents no great problem. . . Men who are active in a fast-moving field know that only a limited number of workers are likely to make significant discoveries, and that they themselves can keep abreast of new developments if they are in close contact with those few."[2]

To these people key words will seem quite useless—until they want some information in an auxiliary field and realize that an up-to-date compilation of magnetic moments, say, was made possible by the key-word system.

7.C | *Nuclear Physics* 56 (1964) 593–603; © *North-Holland Publishing Co., Amsterdam*

Not to be reproduced by photoprint or microfilm without written permission from the publisher

**PERIPHERAL REACTIONS IN COPPER INDUCED BY 19 GeV PROTONS**

G. RUDSTAM

*CERN, Geneva*

Received 12 February 1964

**Abstract:** The reactions $Cu^{65}(p, p\pi^-)Zn^{65}$ and $Cu^{65}(p, p\pi^+)Ni^{65}$ have been studied at 19 GeV. The ratio between the cross-sections is in agreement with the theory developed by Ericson, Selleri and Van de Walle. The experiment thus supports the application of the peripheral interaction model to this kind of nuclear reaction.

E | NUCLEAR REACTION $Cu^{65}(p, p\pi^+)$, $(p, p\pi^-)$, $E_p = 19.2$ GeV; Cu, Zn, Al(p, $xnx'p$), $E_p = 600$ MeV; measured $\sigma$ RADIOACTIVITY $Ni^{57}$, $Zn^{62}$, $Ga^{66}$, $Ga^{67}$ [from spallation]; measured $T_{\frac{1}{2}}$. Natural targets.

**Fig. 1**

*Authors are asked to select from the key-word list the words or phrases appropriate to their articles and to write them on a separate sheet of paper in the style suggested by the sample phrases.*

**KEY-WORD LIST**

Please write phrases containing *all* of the following words or symbols which are appropriate to your paper. After "deduced", please list only quantities which are based on *your* measurements.

**RADIOACTIVITY**

| | |
|---|---|
| Nucleus [from Target (u, v)] | Nucleus |
| measured $T_{\frac{1}{2}}$ (parent) | deduced levels |
| $E_{\alpha,\beta,\gamma,x}$ | $J$ |
| $I_{\alpha,\beta,\gamma,x,ce}$ | $\pi$ |
| sf | $T_{\frac{1}{2}}$ (product) |
| delayed n | $\mu$ |
| $\beta\gamma$-, $\gamma\gamma$-, x$\gamma$-coin (or delay) | $Q_0$ |
| $\beta$ ce-, $\gamma$ ce-, x ce-coin (or delay) | other |
| polarization | |
| cc | |
| $Q_\beta$ | Enriched (or natural) target |
| other | |

**SAMPLE PHRASE**

E | RADIOACTIVITY $^{131}$Te[from $^{130}$Te(n, $\gamma$)]; measured $E_\gamma$, $I_\gamma$, $\beta\gamma$-, $\gamma\gamma$-coin. $^{131}$I deduced levels. Enriched target.

**Abbreviations**

| | | | |
|---|---|---|---|
| cc | internal conversion coefficient or ratio | $Q_0$ | intrinsic electric quadrupole moment |
| ce | conversion electrons | sf | spontaneous fission |
| coin | coincidence study | $T_{\frac{1}{2}}$ | half-life |
| $E$ | energy | x | x-rays |
| $g$ | gyromagnetic ratio ($\mu/J$) | $\alpha$ | alpha particles |
| $I$ | intensity, relative or absolute | $\beta$ | beta particles |
| $J$ | total angular momentum quantum number | $\gamma$ | gamma rays |
| levels | level energies | $\epsilon$ | electron capture |
| $Q$ | (1) electric quadrupole moment | $\mu$ | magnetic dipole moment |
| $Q$ | (2) energy released in ground-state to ground-state reaction | $\pi$ | parity, $\pi$ mesons |
| | | $\sigma$ | cross section |

**Fig. 2**

### *References*


1. F. W. Lancaster, J. Mills, American Documentation **15**, 4 (1964).
2. P. Abelson, Science, July 26, 1963, p. 319.