The Internet as a Source of Sci-tech Information – Cambridge Scientific Abstracts' Experience

Mark Furneaux
Managing Director, European Operations, Cambridge Scientific Abstracts

Introduction

Cambridge Scientific Abstracts was founded in the late 1950's. In 1971, when purchased by Messrs. Hixon and Snyder, it published several print scientific/technical abstracting journals.

Cambridge Scientific Abstracts' initial acquisitions were four titles published by Data Courier, Inc. and seventeen additional scientific titles from Information Retrieval, Inc. Since that time, Cambridge Scientific Abstracts has maintained growth by acquiring new titles, developing new journals, and pioneering methods of electronic information delivery. In 1996, Cambridge Scientific Abstracts acquired Materials Information from ASM International and the Institute of Materials (UK). In March 1998 Cambridge Scientific Abstracts acquired Sociological Abstracts from Sociological Abstracts Inc. and moved into the 'softer' sciences for the first time. Today, Cambridge Scientific Abstracts publishes over 80 print journals and databases in diverse, cutting edge areas of scientific research such as biotechnology, superconductivity, materials science, AIDS, and environmental pollution.

Cambridge Scientific Abstracts has long been a pioneer in creating products for new channels of electronic dissemination of information. In 1985 it was the first publisher to offer MEDLINE from the U. S. National Library of Medicine on CD-ROM. This business was later sold to Silver Platter. Cambridge Scientific Abstracts has a long history of supply data for dissemination on local area networks and wide area networks, and Cambridge Scientific Abstracts data is now available through many remote access hosts.

Cambridge Scientific Abstracts' current initiatives involve maximizing the information distribution opportunities of the Internet. In 1994, Cambridge Scientific Abstracts launched its Internet Database Service, which initially provided parallel electronic database versions of all Cambridge Scientific Abstracts print journals (with the addition of backfile information. For pioneering this effort the Information Industry Association awarded Cambridge Scientific Abstracts with its "Best Science/Technology Database/Service" award for 1994. Now in addition to Cambridge Scientific Abstracts titles, many other publishers' databases are available through the Internet Database Service, with additional publishers' titles being added all the time.

The Internet Source

Today I am going to leave aside the searching of bibliographic databases delivered via the Internet – they represent a huge information resource which I will be talking about tomorrow in my paper „Searching for Sci-Tech Information via CSA's Internet Database Service (IDS)". Today I will focus on our experience of trying to turn the huge, variable and undisciplined resource of information of quite literally every type, we call the Internet into a valuable, searchable source of scientific information.

Cambridge Scientific Abstracts began an experiment to expand the boundaries of Internet publishing in 1994, with the release of theEnvironmental RouteNet.This service, still avialable today, offers subscribers a single Internet point of access to a global array of environmental information resources including both proprietary databases and over 2,000 high quality Internet sites selected and annotated by

Cambridge Scientific Abstracts's editorial expertsEnvironmental RouteNetwas awarded the 1995 Information Industry Award for "Best Web Site / Business."

In addition to abstracts databases and annotated web sites, ERN contains full-text news services, legal, policy, compliance matters, and much much more, all fully searchable using a global search feature. Environmental RouteNet's most recent feature is the Policy and Compliance Forum. The Forum is an interactive medium for the discussion of such critical environmental issues as clean air, clean water, wetlands, biodiversity, population, global warming, and ozone depletion.

There is good news and bad news about the Environmental Routenet experiment. Everyone who has seen it loves it as a source of environmental information. At the same time we have found it to be very difficult concept to explain and one from which it is hard to earn money – there is a strong expectation that this type of information will be free on the Internet. Engineering Information have carried the Routenet concept to its logical conclusion with Ei Village; we prefered to focus more of our efforts on the Internet Database Service and other Internet initiatives, in addition to Ernvironmental Routenet.

The release of Version 3 of the Internet Database Service in October 1997 introduced two new exciting web resources. First, as a derivative of part of Environmental Routenet, we released the Web Resources Databases. Beginning with the environment, the Web Resources Database has recently been expanded to cover three more disciplines: the Aquatic, Materials, and Biological Sciences. The Web Resources Database identifies quality web sites, indexes them for effective retrieval, and delivers the full-text by providing quality checked links. This free resource appears automatically during an IDS search, leaving searchers just a few clicks away from these web resources. The second resource - Link to Holdings - links our databases to the web-based journals holding list of the customer organisation. Users can search IDS and at the same time check that the journal cited is in their library.

The CSA Hot Topics Series is produced as a public service and features reviews of timely subjects. Every Hot Topic includes a mini-review of the scientific literature accompanied by a set of recent key citations with abstracts. Several or all of the following are also included in each Hot Topic: Web sites of interest, pertinent books, review articles, and a glossary. Our "Update on Aging" contains three Hot Topics thus far: Alzheimer's disease, osteoporosis, and the immunology of aging. Our most recent Hot Topic is concerned with the effects of El Niño on marine life.

Some of the issues with Internet sources

The creation of the Web Resources Database illustrates some of the issues surrounding Internet content. Our scientific editors spend a lot of time searching the web for quality sites - the information itself may be free but it is very time consuming and expensive to find the resources in a systematic manner. We use the search engines, we follow hundreds of links, we hear of sites from others, we see sites mentioned in the literature, we find some by luck. We examine them all, we reject most usually for lack of reliable content, and the remainder - usually around 20%, we index at the home page or deeper down content level, and create links to the resource itself. Links break, so every month we check the links using a mixture of computer checking and human intervention on the doubtful addresses. The environmental part was released first, and now contains some 2,000 quality sites, selected from over 10,000 contenders. Even high quality sites vary a lot. Some have literally hundreds of frames and photographs, often adding little to the scientific content of the site – that makes them horribly slow to use, especially in countries with variable connectivity. As an aside that is one reason why we have kept the design of our Internet Database Service web-site short on graphics. Graphics and speed have an inverse relationship to one another.

We welcome all efforts to bring discipline out of chaos on the Internet. Cambridge Scientific Abstracts has made its major contribution, Ei Village is helping engineers world-wide, umbrella organisations like Species 2000 have sprung up to try and focus efforts and funding, in their case to pull together the biosystemtic and nomenclature databases of the world.

In conclusion

Cambridge Scientific Abstracts has, on its own, made a massive contribution to Internet content with the availability of over 70 databases and over 7 million summaries of mainly peer-reviewed scientific and technical articles. The Environmental Routenet and the Web Resources Database have begun to make sense of the maze of free scientific and technical information available on the web. They are beginning to do for the web, for the scientific community at least, what abstracting and indexing services have done so successfully for journals for over a hundred years - select what is valuable, summarise the content, index it for accurate retrieval, and lead the searcher to the original item.