

**AC 2010-536: SEEKING AND FINDING THE AEROSPACE LITERATURE FROM
1996 - 2010: AND, THE WINNER IS . . . GOOGLE**

Larry Thompson, Virginia Tech

Seeking and Finding the Aerospace Literature From 1996 – 2010:

And, The Winner Is Google

Abstract

The *Scientific and Technical Aerospace Reports* (STAR) has been a standard resource in libraries since its inception in 1963. Beginning in 1996 the title was only available online and recently NASA has limited online access to the most recent two years. This paper compares the indexing in STAR with other standard resources such as the NASA Technical Reports Server, the Aerospace and High Technology Index from ProQuest, NTIS, and Google. It finds that Google has the most comprehensive indexing of the STAR entries.

Introduction

In 1963, NASA, through its Office of Scientific and Technical Information, began publishing the *Scientific and Technical Aerospace Reports* (STAR). This abstract journal was a mainstay in engineering libraries for over three decades. In 1996, beginning with Volume 34, NASA ceased production of the hardcopy version of STAR, and continued with an online only version. For several years all volumes of STAR beginning with 1996 were posted and accessible online.

It is now NASA's policy to post online only issues from the current volume, plus the full previous volume. In December, 2009, users could access all issues of Volume 46, (2008), and Volume 47, (2009). In January 2010, NASA removed all issues published in Volume 46, (2008), so that only issues from Volume 47, (2009), and those published in Volume 48, (2010), were available. Thus, the availability of issues varies from one to two years, depending upon the point in the publication year.

The purposes of this research are twofold:

First, to determine if the online NASA Technical Reports Server (NTRS) completely indexes all of the NASA generated papers indexed in STAR. Based upon a paper presented at the 2009 ASEE Conference, "NACA / NASA Document Indexing: 1915 – 1995,"¹ it is expected that the NTRS will duplicate the level of NASA document indexing provided by STAR.

The second purpose is to determine if there is indexing available for the non-NASA generated papers which are contained in STAR. STAR indexes material in a wide variety of aerospace related subject areas, from a number of sources, and it is uncertain how well the non-NASA papers are indexed.

The purpose of the paper is not to determine what resources are best for conducting a comprehensive literature search within the aerospace field. Rather, it is to determine if complete citation information for STAR entries can be found in resources outside of STAR.

Methodology

NASA uses STAR to index material in several subject areas including Aeronautics, Astronautics, Chemistry and Materials, Engineering, Geosciences, Life Sciences, Mathematical and Computer Sciences, Physics, Social and Information Sciences, and Space Sciences. Although the inclusion of the citations in STAR confirms the relevance of all of these areas to aerospace, for the purposes of this paper only those citations in the first two categories, Aeronautics and Astronautics were surveyed.

The first issue of volumes 46, (2008); 44, (2006); 42, (2004); 40, (2002); and 38, (2000) were sampled. At the beginning of the research process, the author was under the false impression that NASA would provide backfiles of STAR issues. However, NASA was unwilling to do this, and stated via e-mail that, "STAR issues from 1996 through 2007 are no longer available from here at the NASA Center for AeroSpace Information. STAR is produced as a current awareness tool and is not considered to have lasting value as a reference tool."² NASA further stated that, "Requests for one or two issues of STAR may be honored on a case-by-case basis, but we do not provide bulk quantities of STAR, whether as individual issues or as volumes."³ Therefore, the author's request for the first issue of each volume was denied by NASA. In order to obtain STAR issues to sample, the author searched the WWW and was able to find institutions that had posted back issues of STAR, which were downloaded.

A total of 100 citations was selected and these citations were searched for in the following databases:

- Aerospace and High Technology Database (AHT) through the ProQuest / CSA interface.
- NTIS, INSPEC, and Compendex (Comp) through the Engineering Village interface.
- Google (not Google Scholar)
- NASA Technical Reports Server (NTRS) at <http://ntrs.nasa.gov/search.jsp>

Searching was done initially through title word phrase searching. If no results or multiple results were found, then the search was expanded or limited by author searching. The purpose of the search was to mimic the situation in which a patron has title and / or author information for a document, but is unable to find a complete citation. It is a common occurrence with both NASA generated and grey aerospace literature that a user will have incomplete citation information.

Survey search results

The following chart summarizes the number of records found in each database for each major document type. The top portion shows "hits" for the NASA generated documents, while the bottom shows the non-NASA documents. Both sections are subdivided to show the major classes of documents contained within each. Column two shows the number of documents in each class, while columns three through eight show the number of times bibliographic information for each document type was found in the indicated resource. The resource columns are ordered by the number of total hits in each resource.

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8
	Number of Documents	Google Hits	AHT Hits	NTIS Hits	NTRS Hits	Comp Hits	INSPEC Hits
<u>NASA Generated Documents</u>							
Technical Document	33	31	23	26	32	2	0
Conference Papers	20	18	15	14	20	0	0
AIAA Papers	7	7	5	3	7	1	0
Misc	4	4	4	0	4	0	0
Sub-total NASA	64	60	47	43	63	3	0
<u>Non-NASA Documents</u>							
Conference Papers	5	5	5	1	0	1	2
DOE	7	7	7	7	0	0	0
DTIC	13	13	6	12	0	0	0
Misc	11	9	10	8	0	1	1
Sub-total Non-NASA	36	34	28	28	0	2	3
Total	100	94	75	71	63	5	3

Discussion of the survey results

With regard to the NASA generated literature, both Google and the NTRS do an excellent job, with NTRS compiling an almost perfect score of 63, while Google is slightly lower at 60. Interestingly, the one document which NTRS misses is covered by Google. So, between the two, they cover 100% of the STAR indexed NASA literature. AHT and NTIS are in a second tier behind NTRS and Google. There is no discernible pattern to the omissions in the NTIS indexing. However, the AHT statistics are skewed by the fact that while it indexes 42 out of 43 NASA documents for years 2000 – 2006, it only indexes 5 out of 21 NASA documents from the 2008 STAR. Thus, the shortfall in AHT indexing may be one of currency rather than comprehensiveness.

Within the non-NASA literature, Google leads the way, with both AHT and NTIS doing a credible job. The lack of any hits in the NTRS for non-NASA literature confirms that it is indeed the "NASA" Technical Reports Server and not a general aerospace literature database.

Looking at the total of all documents, Google is clearly the best resource, indexing 94 out of 100 STAR entries, and showing no weak areas. In five out of the nine document types, Google either indexed the greatest number of papers or tied for the lead. In addition to the 94 hits indicated in the chart, Google also linked to the pdfs of the STAR issues which contained two more items. However, because Google did not give a direct link to the item's bibliographic information, but only linked to the STAR pdf containing the item, these two were not counted in the Google total.

AHT did an excellent job of indexing the literature from the 2000, 2002, 2004, and 2006 volumes of STAR, providing hits on 70 of the 74 STAR entries sampled. If it had continued at that pace with the 2008 STAR it would have been very close to the Google total. However, it only indexed 5 out of 21 NASA documents, and none of the 5 non-NASA documents, from the 2008 STAR, leading to a distant second place finish.

Full text access considerations

A primary consideration when evaluating an indexing resource is the availability of links to the full text of the document indexed. While searching for document indexing in the above resources, a count was also made for links to online full text. In this area Google also prevailed, providing links to the online full text for 66 citations, while NTRS had links to the full text for only 48.

Although AHT gives bibliographic information for 75 documents, it only provides full text links to 6 documents through Serials Solutions. Five of the linked documents in AHT were from AIAA and one was from NASA. AHT therefore lags significantly behind both Google and NTRS.

Similarly, although NTIS indexes a significant number of documents, there were no links to the full text from the database. Most of the NTIS records indicate the document should be ordered from NTIS, even when the full text is freely available online. Many times Google links directly to the full text at the NTRS, DTIC, or DOE site, while NTIS offers only ordering information.

Analysis of materials indexed in STAR

Although NASA was not able to provide the backfiles of the requested STAR issues, it was able to provide excellent data about the source material for entries in STAR⁴. For the twelve year period from 1996 – 2007 a total of 11,544 records appear within the first issue of each year's volume. The sources for these documents are as follows:

- 6,983 (60%) non-NASA records – from the Department of Energy, the Defense Technical Information Center, or the National Technical Information Service.
- 2,042 (18%) NASA records – most likely available on the NASA Technical Reports Server.
- 1,489 (13%) subsidiary records – those without complete source information.
- 1,030 (9%) non-NASA records – from various sources, foreign and U.S., most from sources that contributed in the 1990s but no longer do.

Because the NTRS only indexes NASA generated documents, less than 20% of the total STAR records will be covered in NTRS. The remainder of the STAR documents will need to be searched for in Google, NTIS, or AHT.

An alternative to the surveyed databases

For some researchers, there may be another method of accessing the STAR information. STAR is produced from NASA's primary information product, the NASA Aeronautics and Space Database (NA&SD), which is the source for all information published in current and past issues of STAR.

The NA&SD is available to U.S. Government agency civil servants and their primary contractors. Theoretically, this includes personnel in libraries that serve researchers who have active NASA grants or contracts. However, it appears that access is tied to a specific active grant, and is only in effect for the duration of the grant. Therefore, as one research grant expires and another takes effect, it would be necessary for librarians to reapply for access under the new grant. Also, the database access is tied to a specific IP address, which limits its usefulness in a library setting.

Conclusions

Several conclusions can be drawn from this survey. Many of these were unexpected by the author at the beginning of the research and some are frustrating. None of them present an optimistic picture for aerospace indexing.

Aerospace and High Technology Database - First among the unexpected was the precipitous fall off of recent indexing in the Aerospace and High Technology database from ProQuest CSA. This database has been the author's "go to" resource for all things aerospace for many years. The expectation at the beginning of this research was that AHT would lead all other databases in comprehensiveness, and would prove to be a worthy replacement for STAR. While this was generally true for the STAR volumes of 2000 - 2006, the indexing of only 5 out of 26 STAR records from 2008 leads to many questions, such as: Has there been a change of indexing policy for the AHT? Is the lack of indexing for recent material simply a time lag? If a time lag, is the time lag a one-time fluke, or is a delay of nearly two years the norm for this index?

The author contacted Rich Hummel from ProQuest in December, 2009, with preliminary questions. Rich reported⁵ that about 93,000 records from AIAA/NASA appear in the AHT. Of those 93,000 records, there are about 85,000 STAR records. Of significant interest is the fact that Rich reported that the most recent STAR records in AHT at the time of his e-mail, (December 2009), were from 2008. With a reported coverage lag of one year from ProQuest, and an observed lag of two years from the survey, it does not appear that the AHT is providing the current indexing that researchers require and expect.

It is evident that Google has surpassed the AHT in coverage of the aerospace literature from STAR. It seems that it would be a straightforward matter for ProQuest to import the STAR data into the AHT database, and by doing so provide comprehensive aerospace coverage in a credible search interface. Why ProQuest has not done this is unknown. The author made an attempt in February, 2010, to obtain further information about AHT coverage from ProQuest, but has not received a reply.

Google as the go-to index – The second unexpected result, which mirrored the disappointment in the AHT, was the surprise in Google's performance. It has become increasingly evident over the past few years that Google is a legitimate search tool for many situations. Still, the author was startled to see how it clearly outranked established databases, both in the ability to retrieve bibliographic records, and to link to the online full text. Of the 100 STAR citations searched, Google indexed 94 and linked to the full text of 66. The next highest products indexed only 75 (AHT) and linked to the full text of 48 (NTRS). It is noteworthy that Google placed first in both indexing and full text linking, while the second place tallies were split between two products. Google appears to be the one-stop-shop.

Confusion about aerospace indexing within the library community – Third among the unexpected findings was the confusion within the library community about the availability of STAR volumes online. While searching for full text STAR documents on the WWW, the author found many references to STAR on library websites. A number of these sites stated that NASA provided online access to STAR volumes from 1996 to the present, while in fact NASA's policy of only providing the most recent two years has been in place for at least three years.

NASA's narrow view of aerospace literature indexing – While the above findings were unexpected, the most frustrating aspect of the research was the NASA policies which sequester the STAR material and make it cumbersome to use by researchers. This has been done in three ways:

- NASA has decided to make older volumes of STAR unavailable online. The NASA justification for this policy is that STAR is a current information product, and is not intended for long term searching. Even if this assumption is true, which is not certain, it is still unclear why NASA would decide to remove online access for the STAR volumes. In an era of cheap online storage and increasingly fast internet access, it makes no sense. Why not leave the volumes online, and allow researchers to decide how the information will be used?
- NASA has decided not to distribute the information via DVD or other media. Not only has NASA removed the data from online, but it also refuses to release the data in another format. Why has it done this?
- NASA has decided to issue the online data in the pdf format, which is difficult to search. This is a disservice to researchers and places NASA outside the norm of other government supported information agencies. The National Library of Medicine (Medline), the National Technical Information Service (NTIS), the Transportation Research Board (TRIS), and the National Agricultural Library (Agricola), have all placed their information indexes online in a freely searchable format. In contrast to this, NASA has chosen to display its data in pdf. Why can't NASA follow the lead of other government agencies and provide the info in a readily searchable format?

None of these decisions by NASA make sense to the author. In an era when other indexes and journals are steadily increasing the availability of their online backfiles, NASA has chosen to limit its online availability. The expenditure for gathering and compiling the data has already been made. Why not maximize the usefulness of the data by making it readily available?

The void in quality aerospace indexing - Based on the above findings, it seems that there is no excellent way to search the aerospace literature. The AHT database is obviously not current with regard to much of the literature indexed in STAR. The online STAR volumes are in pdf format and are not easily searched. Google, while it covers much of the literature, does not formally index the material and cannot be regarded as a suitable search engine for a comprehensive

subject based search. Thus, aerospace has no index which comprehensively covers the literature in the field.

There are two obvious solutions to remedy the situation. The first would be for NASA to follow the lead of other government agencies and make the STAR data freely available and readily searchable for researchers. The second would be for AHT to begin comprehensive and timely indexing of the STAR material, and fulfill its role as the go-to database for aerospace.

1. Thompson, Larry. "NACA / NASA Document Indexing: 1915-1995" in the *Proceedings of the 2009 ASEE Annual Conference & Exposition*. June 14-17, 2009, Austin, Texas. <http://soa.asee.org/paper/conference/paper-view.cfm?id=10843>
2. Paulson, Edna. NASA STI Help Desk. E-mail. December 11, 2009. edna.w.paulson@nasa.gov
3. Ibid.
4. NASA Help Desk. E-mail. November 25, 2009. helpdesk@sti.nasa.gov
5. Hummel, Rich. Senior Product Manager, ProQuest. E-mail. December 21, 2009.