
STINFO COPY

AFRL-HE-WP-TR-2007-0079

# Engineering Data Compendium: Human Perception and Performance Conversion to HTML

**Patrick K. Clark**

**Northrop Grumman Information Technology**
**2555 University Blvd.**
**Fairborn OH 45324**

**November 2006**
**Final Report for July 2006 to November 2006**



**Air Force Research Laboratory**
**Human Effectiveness Directorate**
**Warfighter Interface Division**
**System Control Interfaces Branch**
**Wright Patterson AFB OH 45433**

# NOTICE AND SIGNATURE PAGE



**TECHNICAL REVIEW AND APPROVAL**

**AFRL-HE-WP-TR-2007-0079**

**THIS TECHNICAL REPORT HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION.**

**FOR THE DIRECTOR**

| | |
|---|---|
| **//signed//** | **//signed//** |
| Kristen K. Liggett, Ph.D. | Daniel G. Goddard |
| Program Manager | Chief, Warfighter Interfaces Division |
| System Control Interfaces Branch | Human Effectiveness Directorate |

# REPORT DOCUMENTATION PAGE

*Form Approved*
**OMB No. 0704-0188**



| 1. REPORT DATE *(DD-MM-YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From - To)* |
|---|---|---|
| 29-11-2006 | Final | Jul 2006 – Nov 2006 |

**4. TITLE AND SUBTITLE**
Engineering Data Compendium: Human Perception and Performance Conversion to HTML

**5a. CONTRACT NUMBER**
FA8650-05-D-6633

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**
62202F

**6. AUTHOR(S)**
Patrick K. Clark

**5d. PROJECT NUMBER**
7184

**5e. TASK NUMBER**
09

**5f. WORK UNIT NUMBER**
T4

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Northrop Grumman Information Technology
2555 University Blvd
Fairborn OH 45324

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Air Force Materiel Command
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Interface Division
System Control Interface Branch
Wright-Patterson AFB OH 45433-7511

**10. SPONSOR/MONITOR'S ACRONYM(S)**
AFRL/HECI

**11. SPONSORING/MONITORING AGENCY REPORT NUMBER**
AFRL-HE-WP-TR-2007-0079

**12. DISTRIBUTION AVAILABILITY STATEMENT**
Approved for public release; distribution is unlimited.

**13. SUPPLEMENTARY NOTES**
AFRL/PA Cleared 8/2/2007; AFRL/WS-07-1759.

**14. ABSTRACT**
The objective of this task was to convert the three volume Engineering Data Compendium: Human Perception and Performance to HTML so it could be access via the world wide web. This document provides an extraordinary body of knowledge that can be of use to researchers and practitioners in the fields concerning Human System Integration (HSI).

**15. SUBJECT TERMS**
Human perception, human performance, engineering data compendium

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Kristen K. Liggett |
| UNCL | UNCL | UNCL | SAR | 16 | 19b. TELEPONE NUMBER (*Include area code*) |

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI-Std Z39-18

**THIS PAGE INTENTIONALLY LEFT BLANK**

Table of Contents

List of Figures

## 1. Objective

The objective of this task was to create a web accessible version of the *Engineering Data Compendium (EDC): Human Performance and Perception.* This document provides an extraordinary body of knowledge that can be of use to researchers and practitioners in the fields concerning Human System Integration (HSI). The EDC has been published previously in traditional paper format, and also as a Macintosh software application (CASHE:PVS). To allow a much wider audience to access this extensive work, this task will create an HTML presentation of the EDC.

## 2. Background

The EDC currently exists electronically as CASHE:PVS. The technical content is contained in a large number of files. There is a large group of SGML files that contain the technical entries as well as the table of contents, index, a series of Design Questions, a glossary of terms, and data type definitions that codify the structure of each SGML file type. There are also large number of binary files contain the figures and equations used in the EDC. Together this set of CASHE:PVS files formed the set of input data that we were able to use on this task. The strategy for this project was to convert the input source data files into a set of HTML files suitable for presentation on the World Wide Web. To the extent that we succeeded, we were fortunate for the foresight that produced the SGML files for CASHE:PVS, which have the content and structure definition that made automated conversion possible.

## 3. Tasks Performed

In accordance with the Statement of work for this effort, two major tasks were performed, titled **Task 1 – Requirements Gathering and Design Consultation** and **Task 2 – EDC Implementation as an HTML Document Series**. Task 1 involved a detailed analysis of

the existing CASHE:PVS DTDs and other files, as well as a comparison with the paper EDC product. From this it was determined that there was an excellent structure available from which to proceed with the conversion effort. Task 2 involved the detailed conversion of the document with many specific subtasks listed below:

a. Build or modify a limited SGML to HTML translator to cover the CASHE:PVS DTDs.
b. Translate SGML pages from the CASHE:PVS software to HTML
c. Convert single figures from CASHE:PVS software to JPEG format for web display
d. Convert multiple frame figures from CASHE:PVS software to single figures by scanning original figures from printed EDC.
e. Modify or redraw "problem" figures.
f. Renumber figures as necessary.
g. Insert image tags, tables, equations, and captions in HTML
h. Convert tables from CASHE:PVS SGML files to HTML.
i. Convert table of contents, index, and glossary from CASHE:PVS SGML files to HTML.
j. Insert image/table tags for unreferenced figures and tables.
k. Extract equations from Mac resource files.
l. Convert CACHE:PVS Design Questions SGML files to HTML/Javascript.
m. Review generated document.

Rather than address each of these tasks in the order specified in the SOW, they will be addressed in the order in which they were accomplished during the project.

### 3.1. Conversion of Figures

The figures in the EDC are a very important part of the document that convey a large amount of information. Immediately after assessing that the conversion to HTML was possible given the existing DTDs, the figures were looked at to make sure this critical part of the information could be recovered. Since the graphic data was not itself in SGML, we were concerned about how easy it would be to get the figures into a web accessible format. They existed with the CASHE:PVS files as Macintosh

PICT files. Fortunately, there is a very good program on the Macintosh for converting between graphic file formats called GraphicConverter. This resolved most of our concerns and allowed a quick conversion of all the existing graphics to JPEG formatted files. A small amount of effort was required to rename the files. Of more concern was that occurrence of about 80 figures that we called "multi-frame" figures, that is they contained several different pictures which lay on top of one another in their JPEG converted form. Figure 1 shows an example taken from EDC Entry 1.210.

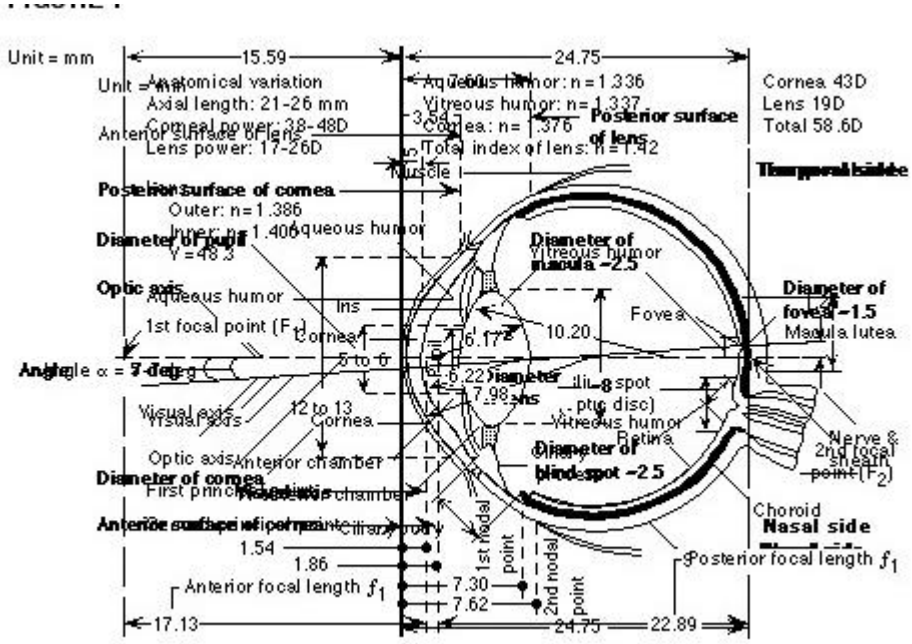



**Figure 1. "Multi-Frame" figure as converted by GraphicConverter.**

This file format allowed CASHE to have several frames for a figure that could be traversed by clicking on radio buttons, "next" buttons, etc. Within the scope of this effort it was not feasible to re-implement this feature within a web browser environment, so the decision was made instead to scan these figures from the paper document and insert those as replacement figures in the EDC HTML version. (As an example, Figure 2 shows the scanned figure that replaced Figure 1.)

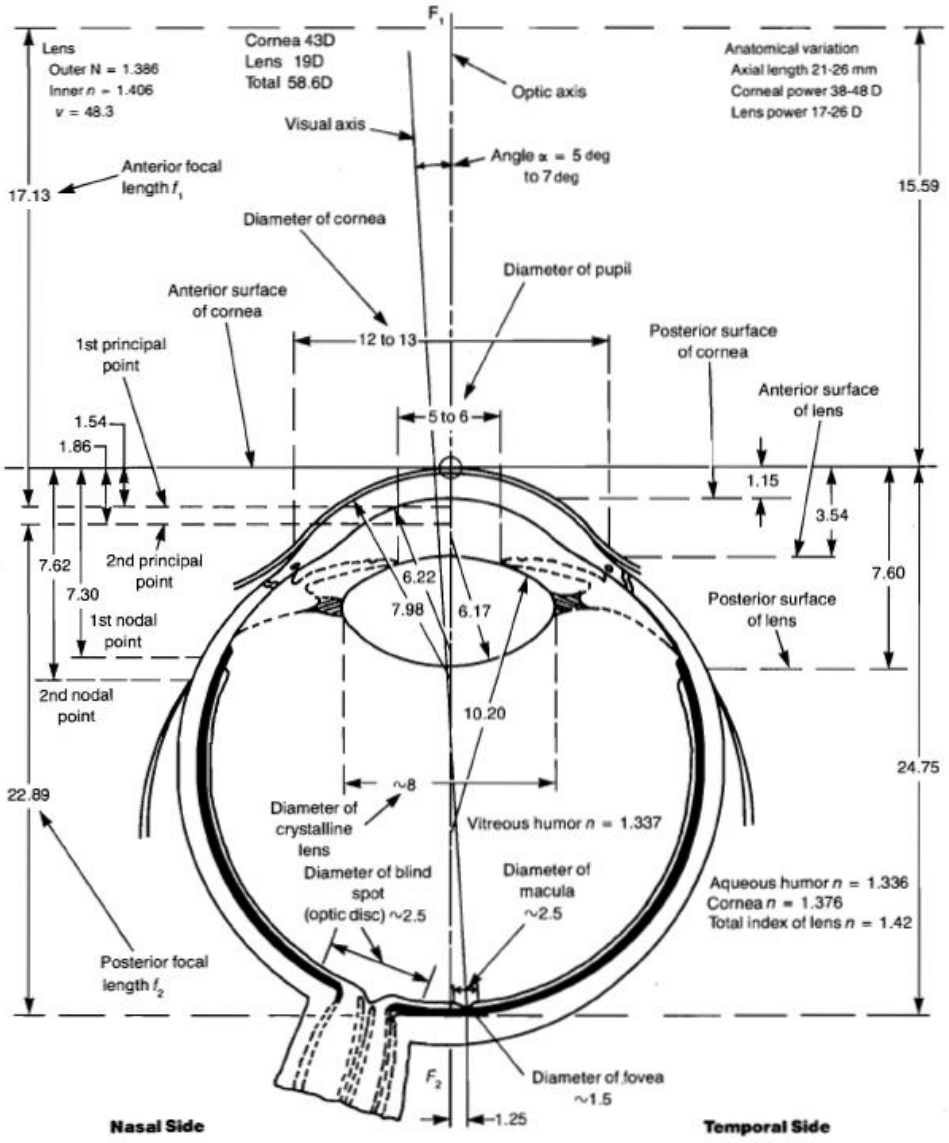



**Figure 2. Replacement scanned graphic for Figure 1.**

Figures are stored in the EDC HTML structure in a figures subfolder within the folder for each Section's entries. For example, entries in Section 1 are in the folder EDCSec01, and the figures for those entries are in a subfolder titled EDCSec01Figs.

### 3.2 Extraction of Equations

Equations are embedded in many of the entries within the EDC, and form a key part of the content of these entries. After some investigation, it was also found that they were in PICT format, but were embedded as resources in Macintosh resource fork file components. After several attempts at extracting these graphics, we found that the GraphicConverter program was again the best means of doing so. In most cases the extraction the program provided was very satisfactory, although in a few cases (notably a few tables) the quality of the graphic seemed less than in looking at the similar resource via the CASHE:PVS program. These files are stored in the HTML in the "resources" directory, under the applicable subfolder (for example, equations and other graphic resources for Section 1 of the EDC are under the folder EDCSec01).

### 3.3 Build SGML to HTML Translator

All of the additional subtasks of Task 2 were incorporated into this subtask, in order to make it as easy as possible to regenerate the complete set of web pages should some global changes be required/desired. Some of the subtasks were actually not required based on the decision to scan the multi-frame figures (modifying the multi-frame figures and renumbering).

Because SGML in the form used in the CASHE:PVS files is very similar to XML, and there are free parsers available for XML, the decision was made to perform a two stage conversion. The SGML files are first converted to legal XML, and the XML is then converted to HTML for web presentation.

The conversion from SGML to XML was not a major effort, in general it involved the following:

- Adding matching end tags for SGML tags that did not have them (XML requires this).

- Making the case of tags consistent (XML tags are case dependent, <xref> is actually interpreted as a different tag than <XREF>)
- Processing of special characters such as Greek letters, foreign language letters, mathematical symbols, etc.. Since the Apache Xerces XML parser we started with does not process Unicode characters, we instead inserted them inside a custom XML tag <SPCHAR>, and then during the second phase of conversion converted them back to either Unicode or mnemonic special characters (e.g., &#x3405, &theta) in the HTML pages.
- Fixing SGML errors. In a small number of cases there are mismatched tags in the CASHE:PVS SGML. Sometimes this was automated, but in some cases these problems had to be fixed by hand. A list can be generated of these changes by running a folder differencing program of the original CASHE:PVS files against the converter input files (Araxis Merge Pro is a good, inexpensive product that can be used for this purpose).

## 4. Results

We have produced an automated conversion tool for the translation of CASHE:PVS files to HTML files which can be easily accessed from a web site or hard disk. The automated generation feature enables global format changes to be easily made to all entries of the EDC should it be required, without editing of all individual entries.

### 4.1 EDC HTML File Structure

The EDC HTML files are organized by section, as shown in Figure 3.

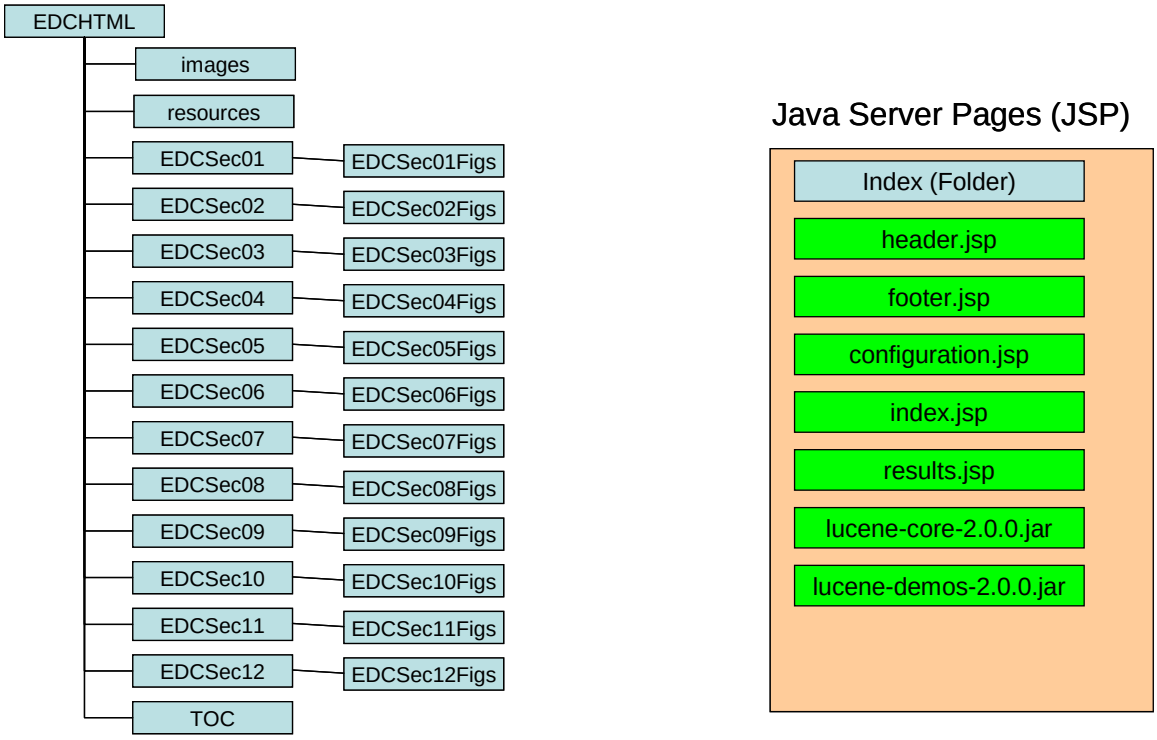
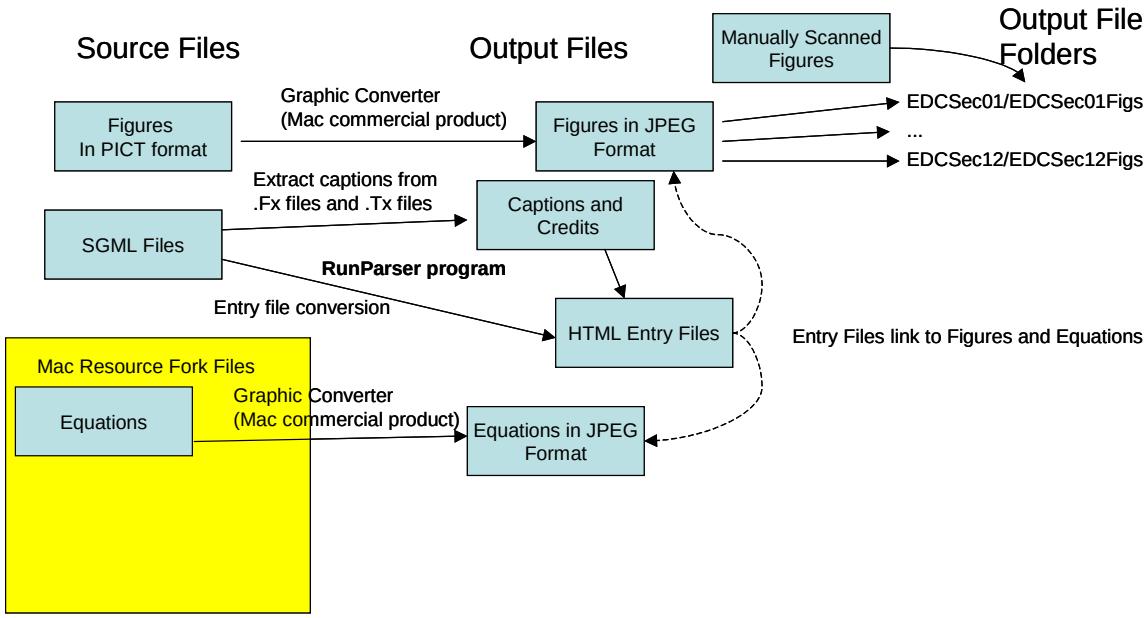



**Figure 3. EDC HTML folder structure.**

Folders titled EDCSecXX contain HTML files for entries in the XX Section of the EDC, as well as a folder called EDCSecXXFigs, which contains the graphics for those entries. The resources folder contains 12 subfolders entitled EDCSecXX which contain the equations for each section (sections 6 and 12 have no equations, so those folders are empty. The images folder contains globally used bitmaps, including folder icons, button bars, the HSIIAC logo, etc. In addition to the static HTML representing the HSIIAC entries, a search function was implemented using Java and Java Server Pages. This functionality could easily be implemented (and has been by others) in many other languages. The choice of Java and JSP for this project was based on the preference of the web administrator of the HSIIAC web page (Alion Sciences, Inc.). In a typical web installation, the static content and the JSP are stored in different folders, so the EDC web reference can be easily configured to accommodate any changes required should the setup change as the HSIIAC web site evolves, or if the EDC is eventually hosted on other sites to improve its accessibility. This configuration is discussed in the following section.

## 4.2 Generation of the EDC Web Reference from CASHE:PVS Files

Figure 4 summarizes the major steps performed during the conversion from the CASHE:PVS Macintosh files to the EDC Web Reference HTML files. The graphic conversion steps shown in the figure were done with the aid of a commercial tool




(GraphicConverter).

**Figure 4. Steps in the conversion of CASHE:PVS files to EDC Web Reference HTML.**

Most of these steps do not need to be accomplished again if there is a need to regenerated the EDC Web Reference. In order to regenerate it, the following steps are required (this process primarily involves setup of an initial folder and running the RunParser program). Start using the delivered CDRL A008, Computer Software Product End Item.

1. Create the "skeleton" output directory.
    a. Make a copy of the folder named "outputSkeleton", then rename it "output".

    b. This contains the figures, images, equations, and a small number of Javascript and CSS files.
2. Unzip the CASHE source files to C:\HEC_EDC.
3. Run the Java class “RunParser”.
    a. i.e. “java –Xm512m RunParser”,. This assumes CLASSPATH variable is set correctly.
    b. A batch file called GenerateEDC.bat is available which should serve to launch RunParser by double-clicking on it in Windows Explorer.
4. RunParser will pop up a dialog box, a file chooser dialog.
    a. This is where you select the CASHE files to be processed into HTML.
    b. If you select folders, all SGML CASHE files in those folders will be processed.
        i. If you also want to process figure captions and tables, check the box that indicates this choice. If you haven’t made any code changes that would affect these, you can save time by leaving it unchecked. If you want a full generation of the web site HTML, go ahead and check it (it only takes about 15 minutes to generate all the HTML for the whole web site).
5. There are 12 sections that need to be selected, plus the TOC folder.
    a. The folders have their names from the CASHE CD, i.e. “EDC Sec01 folder”. (The TOC folder is just called TOC).
    b. The TOC files have to be processed separately at this time, you can’t process the whole folder (it contains files for the electronic version of MIL-STD 1472D which aren’t recognized by the RunParser program).
        i. Select only the files in the TOC folder that are EDC files. These include files that start with “EDC BOB Index”, “EDC DQ”, and “EDC TOC” and “EDC Glossary”.
6. When you click “Open” on the file dialog, RunParser will start processing the selected files/folders. They will show up in the EDCSecXX folders and the TOC folder (where XX is the EDC section number).

a. They will be named by their entry number, for example “e01-0101.html” is entry 1.101. A figure will be something like “e01-0101f1.html” for Figure 1, and “e01-0104t1.html” for table 1 in entry 1.104. Figure and table files are the ones that are linked to the drop down menus on the entry web page – they are the pop up windows for those figures and tables.

7. If you take all the files/folders in the output directory and copy them to a directory on the web site called “EDCHTML”, the EDC will be available on that web site. All paths are relative, so you can put the document tree where you want it.
8. Two file changes are required to enable the search function, since the search function is implemented with servlets and JSP pages. Static HTML and JSP are typically put into two different locations on web sites.
    a. The JSP code needs to know where the static HTML pages are so that it can point its search results to the right entries.
        i. This path is called “webroot” and is currently set by default for the HSIIAC web site to “/products/compendium” in the file configuration.jsp. The EDCHTML directory should be in this webroot directory.
        ii. Configuration.jsp also contains the search index location, which by default is set to “webapps/hsi/compendium/index”.
    b. The static pages have to know how to access the search function
        i. This path is in search.html in the TOC folder – default for current HSIIAC web site is “/hsi/compendium”.

### 4.3 Functions of the RunParser Java Program

The RunParser program is a Java program written by Northrop Grumman for this project, and accounts for the majority of effort expended. For regeneration of the EDC Web Reference, in general all that is needed is to rerun the RunParser program on the CASHE:PVS SGML files. The 12 sections of the EDC can be run in one pass, and the

various TOC files can be run in a second run. RunParser places the resulting HTML files in the directories as shown in Figure 3. Within RunParser, there are two major functions: 1) converting the SGML to valid XML, and 2) converting the XML to HTML for web browser interpretation. The specific major functions it performs are:

•Convert SGML to valid XML

–XML parsers are free and standard!

–Involved "fixup" of tags, since CASHE SGML files had many tags that didn't require end tags

•Conversion of XML to HTML

–Handling of special characters

–Design/implementation of "tree" index structures for TOC,index, Glossary, Design Questions

–Formatting of Tables

–Popups for Figures and Tables

–Bulleted Lists, numbered lists

–Text Emphasis (bold, italic, underline, etc.)

–Links to references, other entries.

–Previous/Next Navigation

•Auxiliary program to crop top Figure labels from all figures

## 5. Conclusions

The EDC Web Reference is a reasonably accurate conversion of the Macintosh hosted CASHE:PVS version of the original Engineering Data Compendium: Human Perception and Performance. Certain aspects of CASHE:PVS unfortunately did not lend themselves easily to automated conversion, and therefore could not be implemented under this project's limited scope and duration. These features mainly included the interactive capabilities and animations implemented for the Macintosh. Still, the information contained in the EDC Web Reference represents the bulk of the

CASHE:PVS content, and most all of the original published paper compendium. A key aspect that enabled the conversion to be accomplished at a relatively low cost was the decision by the CASHE:PVS design team to implement it using SGML. This decision left a legacy of electronic data that could be converted from a standard format in an automated fashion. Had the decision to use a proprietary Macintosh format been made, the conversion would have been much more costly, possibly prohibitively so. It is largely due to this decision that we have been able to make the EDC available to an even wider audience as a web reference on the Internet.