



DEPARTMENT OF THE NAVY  
NAVAL UNDERSEA WARFARE CENTER  
DIVISION NEWPORT  
OFFICE OF COUNSEL  
PHONE: 401 832-3653  
FAX: 401 832-4432  
DSN: 432-3653



Navy Case No. 82714  
Date: 10 October 2008

The below identified patent application is available for licensing. Requests for information should be addressed to:

TECHNOLOGY PARTNERSHIP ENTERPRISE OFFICE  
NAVAL UNDERSEA WARFARE CENTER  
1176 HOWELL ST.  
CODE 07TP, BLDG. 990  
NEWPORT, RI 02841

Serial Number 12/284,478  
Filing Date 19 September 2008  
Inventor Marcus A. Lepore

Address any questions concerning this matter to the Office of Technology Transfer at (401) 832-1511.

DISTRIBUTION STATEMENT  
Approved for Public Release  
Distribution is unlimited

**20081027290**

SGML/XML TO HTML CONVERSION SYSTEM  
AND METHOD FOR FRAME-BASED VIEWER

STATEMENT OF GOVERNMENT INTEREST

[0001] The invention described herein may be manufactured and used by or for the Government of the United States of America for governmental purposes without the payment of any royalties thereon or therefore.

BACKGROUND OF THE INVENTION

(1) Field of the Invention

[0002] The present invention relates generally to conversion of Standard Generalized Markup Language (SGML) and/or Extensible Markup Language (XML) to Hypertext Markup Language (HTML) and more specifically to enhanced functionality of the resulting conversion.

(2) Description of the Prior Art

[0003] Programs to convert SGML and XML to HTML are disclosed in the prior art patents discussed below. As well, the Advanced Interactive Management Technology Center (AIMTC) of the U.S. Navy has designed a software system that provides conversion of SGML

and XML to HTML that has resulted in reducing acquisition and life cycle maintenance costs for Fleet training documentation (e.g., Technical Manuals, Operational Guidelines, etc.) The system manages Standardized General Mark-Up Language (SGML) and Extensible Mark-up Language (XML) objects and then migrates them to Hypertext Mark-up Language (HTML) readable by the Microsoft Internet Explorer (MSIE) 4.0 Service Release (SR-1 or greater) browser. The originally designed solution has numerous useful features. Database Management of SGML components enables SGML components to be compiled and used in multiple locations. The life cycle maintenance costs are reduced because these SGML components are managed as unique objects in the database where they are singularly stored for alteration and used many times in the final display. Use of a free browser as compared to the traditional use of proprietary (for fee) browsers ultimately reduces distribution costs. The use of approved, generic Navy data formats (SGML, PDF, etc.) enables other Navy activities to reuse the source content.

[0004] While present systems are highly useful, the inventors believe it would be desirable to improve the functionality of the resulting converted documents to provide features such as mapping of content classification, multidimensional table of contents, bookmark features, annotation features, local and global search

features and/or other highly desirable functionalities in the manner taught hereinafter.

[0005] Previous efforts to solve problems related to the above are described by the following patents:

[0006] U.S. Patent No. 5,530,852, issued June 25, 1996, to Meske, Jr. et al., discloses methods for extracting profiles and topics from a first file written in a first markup language and generating files in different markup languages containing the profiles and topics for use in accessing data described by the profiles and topics.

[0007] U.S. Patent No. 5,745,360, issued April 28, 1998, to Leone et al., discloses a methodology used to convert non-HTML softcopy documents to HTML formatted documents based on examination of hypertext link entry points. To resolve these links, the invention interrogates the non-HTML formatted texts for hypertext links, verifies that the links are still valid (determines whether the to-be-linked-to topic, document or reference still exists) and constructs HTML link anchors. After all non-HTML links are resolved (i.e., converted to HTML, verified, and HTML anchors constructed), the entire topic or accessed portion of a requested document or topic text is formatted into HTML text and hypertext links. The most common entry point for beginning the conversion process is the table of contents. This usually provides numerous hypertext entry points

to other topics contained in the work. Links can exist at any spot in the work. They are not limited to topic links only.

[0008] U.S Patent No. 5,819,302, issued October 6, 1998, to J. Nielsen, discloses a method and apparatus for automatically converting documents from a first hypertext format that supports multi-layered backgrounds to a second hypertext format that does not support multi-layered backgrounds, such as HyperText Markup Language (HTML). According to the method, a target file is generated that stores all non-background elements of the document in the second hypertext format. A mechanism that reads the first hypertext format is used to display a current page of the document. A screen dump is made of the displayed page after removing or hiding the non-background elements. This process is repeated for all of the pages of the document. References to the graphics files generated by the screen dumps are embedded in the target file. The references cause the background elements of a page to be displayed behind non-background elements of the page when the document is displayed based upon the target file by a mechanism that reads the second hypertext format.

[0009] U.S. Patent No. 6,009,436, issued December 28, 1999, to Motoyama et al., discloses a method, apparatus, and computer program product for mapping a first structured information format to a second structured information format, which allows a user to interactively define the mapping. This invention operates as a

user tool by accepting interactive input from a user of a source input, by processing the input to display the source input in a format for accepting and processing user commands to create or edit a transformation map of source components to target components. Interactive user input is then accepted and processed for selection of an input file to be transformed and selection of a transformation map to be used for the requested transformation. Interactive user input is accepted and processed for selection of individual components of the first structured information format for mapping, and for selection of options for the target components. Exemplary options for the target components are a null value, the source component itself, a single selected target component, or plural selected target components. Interactive user input is accepted for processing to assign attribute values to components of the second structured information format. Exemplary options for the sources of attribute values are attribute values obtained from the source components, system attribute values, no value, attribute values input interactively by the user, and content of element. Interactive user input is then accepted and processed to initiate processing of a transformation of the source input file in the first structured information format to a target output file in the second structured information format.

[0010] U.S. Patent No 6,202,072, issued March 13, 2001, to A. Kuwahara, discloses an apparatus for processing standard generalized markup language document generates a SGML conversion form file correlating a prototype file having a specific form with document type definition by a SGML conversion form generation module, and converts, in a case where a plain text document prepared using the specific form is converted to a SGML document, the prepared plain text document to the SGML document by referring to the SGML conversion form file in a SGML document generation module, while in a case the reverse conversion is executed, the converted SGML document is reversely converted to the plain text document by referring to the SGML conversion form file in a SGML document read-in module.

[0011] US Patent No 6,393,442, issued May 21, 2002, to Cromarty et al., discloses a method and system for converting a source document into a plurality of documents, each of the plurality of documents having one of a plurality of formats. The method and system comprise providing a document type definition for formatting the source document. The method and system further comprise providing a transform to convert the source document into the plurality of documents. At least one of the plurality of documents has a binary code format. The method and system enable production of the plurality of documents, each

representing a version of the source document, such that the versions are consistent with each other.

[0012] U.S Publication No. 2002/0122060, published September 5, 2002, to S. O. Markel, discloses a system for generating HTML code for various HTML interfaces such as various set-top boxes using non-technical personnel. The system employs a graphic user interface such as wizard to allow a non-technical user to enter content data to be displayed on the HTML page. The content data is translated by an XML translator into XML code. An XSL parser for a particular HTML interface is used to translate the XML code into HTML code for each particular HTML interface.

[0013] US Publication No. 2002/0194227, published December 19, 2002, to Day et al., discloses an adaptive transformation and User Interface system that enables transformation of a file or document (e.g., an SGML, XML, HTML or other multimedia file or document) from one format to another format. The transformation supports error correction, filtering and collation of elements of a source document for output and is performed in response to control information comprising transformation parameters. The system transforms a document encoded in a language including presentation style determination attributes from a first format to a different second format. The system includes a source of transformation parameters determining a desired presentation style and content structure as well as an input document

processor. The input document processor transforms a received input document in a first format by parsing the input document and collating elements of the input document into a hierarchically ordered structure representing an intermediate document structure. The system employs a transformation processor for transforming the intermediate document structure into an output document with the desired presentation style of a second format in response to the transformation parameters.

[0014] The above listed patents do not provide for expanded functionality of the converted document such as mapping content classification, multidimensional table of contents, bookmark features, annotation features, local and global search features and other highly desirable functionality as taught herein. Consequently, those skilled in the art will appreciate the present invention that addresses the above and other problems.

#### SUMMARY OF THE INVENTION

[0015] An object of the present invention is an improved method of taking Standardized General Mark-up Language (SGML) data and its logical equivalent Extensible Mark-up Language (XML), formatted in compliance with predetermined rules (this can be style guidance such as the NAVSEA02 Rev D DTD) and converting it

into HyperText Mark-Up Language (HTML) suitable for viewing using the Microsoft Internet Explorer (MSIE) 4.01 SR1 or greater web browser.

[0016] These and other objects, features, and advantages of the present invention will become apparent from the drawings, the descriptions given herein, and the appended claims. However, it will be understood that above listed objects and advantages of the invention are intended only as an aid in understanding aspects of the invention, are not intended to limit the invention in any way, and do not form a comprehensive list of objects, features, and advantages.

[0017] Accordingly, the present invention provides a method for converting SGML or XML source documents to HTML for viewing the SGML source documents in a web browser which method may comprise one or more steps such as, for instance, converting each tag, entity, and attribute of a source document to HTML output files such that each chapter, index, appendix, and glossary are divided into separate HTML output files, creating a plurality of browser frames for viewing the HTML output files comprising at least a content web browser frame, a local navigation web browser frame, and a global navigation web browser frame. Other steps may comprise providing that the content frame comprised of the HTML output files combined with graphic images from the source document for the content web browser frame, and/or generating a

table of contents for the local navigation frame, and/or providing that the content frame is selectively scrollable whereby when the content frame is scrolled then the table of contents is also scrolled to include a portion of the table of contents related to the HTML output files and graphic images displayed in the content frame. The method may further comprise utilizing a Java or JavaScript executable for the steps of converting, creating, and generating.

[0018] The step of generating the table of contents may further comprise mapping chapter, section, paragraphs, subparagraphs, figures, foldouts, tables, multimedia components from the source document to thereby provide an expandable/collapsible table of contents. The method may further comprise creating a classification browser frame based on an attribute of the source document relating to a classification of the source document, e.g., "Top Secret".

[0019] The method may further comprise parsing the source document and stripping out structure elements to thereby generate a plain text file and/or producing a local query interface to interactively query for matches with the plain text file to provide local searching capability. To provide global searching, the method may further comprise parsing a plurality of documents related to the source document and stripping out the structures elements to generate one or more plain text files and/or

producing a global query interface to interactively query for matches with the one or more plain text files to provide global search capability.

[0020] Other steps may comprise mapping an attribute defined in the source document to a bookmark icon displayed in the web browser and/or providing a Java or JavaScript executable application code that when enabled by selection of the annotation icon allows a user to interactively query for bookmarks and/or mapping an attribute defined in the source document to an annotation icon displayed in the web browser and/or providing a Java or JavaScript executable application code that when enabled by selection of the annotation icon allows a user to interactively produce and retrieve annotations.

[0021] The method may further comprise providing HTML hyperlinks for all SGML tags and attributes may comprise xref, xerfid, glossary, and term to provide glossary, term, and definition links viewable in the web browser.

[0022] Accordingly a software system is also provided for conversion of SGML or XML source documents to HTML such that a user can view the SGML or XML source documents in a web browser. The software system may comprise one or more elements such as, for instance, an executable code operable to convert each tag, entity, and attribute of a source document to HTML output files such that each chapter, index, appendix, and glossary is divided

into separate of the HTML output file. The executable code is preferably operable for creating a plurality of browser frames for viewing the HTML output files may comprise at least a content frame, a local navigation frame, and a global navigation frame. The content frame is preferably comprised of the HTML output files combined with graphic images from the source document for the content frame. The content frame is preferably navigable such as with an associated scroll tag. The local navigation frame is preferably generated by the executable code to provide a table of contents that is navigable with an associated scroll tag. In one preferred embodiment, the executable code comprises Java or JavaScript.

**[0023]** The software system executable code is preferably operable for generating the table of contents by mapping chapter, section, paragraphs, subparagraphs, figures, foldouts, tables, multimedia components from the source document to thereby provide an expandable/collapsible table of contents.

**[0024]** The software system executable code is, in one embodiment, operable for creating a classification browser frame based on an attribute of the source document relating to a classification of the source document.

**[0025]** The system may further comprise a parser operable for parsing the source document and stripping out structure elements to thereby generate a plain text file and/or a local query

interface to interactively query for matches with the plain text file to thereby provide local searching capability. The system may further comprise a parser for parsing a plurality of documents related to the source document and stripping out the structures elements to generate one or more plain text files and/or a global query interface to interactively query for matches with the one or more plain text files to provide global search capability.

[0026] In a preferred embodiment, the system is operable for mapping an attribute defined in the source document to a bookmark icon displayed in the web browser and providing a Java or JavaScript executable application code that when enabled by selection of the annotation icon allows a user to interactively query for bookmarks. Additionally, the system is operable for mapping an attribute defined in the source document to an annotation icon displayed in the web browser and/or providing a Java or JavaScript executable application code that when enabled by selection of the annotation icon allows a user to interactively produce and retrieve annotations. Preferably, the system executable code is operable to provide Glossary, term, and definition links viewable in the web browser.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0027] A more complete understanding of the invention and many of the attendant advantages thereto will be readily appreciated as the same becomes better understood by reference to the following detailed description when considered in conjunction with the accompanying drawing, wherein like reference numerals refer to like parts and wherein:

[0028] FIG. 1 is a table which shows an SGML to HTML tag map for an original Java filter in accord with the invention;

[0029] FIG. 2 is a table which shows an SGML/XML attribute map for use in accord with the present invention;

[0030] FIG. 3 is a web browser view of sample SGML/XML tags and attributes such as those identified in FIG. 1 and FIG. 2 in accord with the present invention; and

[0031] FIG. 4 is a web browser view of a sample HTML display produced from SGML/XML tags and attributes identified in FIG. 1 and FIG. 2 in accord with the present invention.

## DETAILED DESCRIPTION OF THE INVENTION

[0032] The present invention provides for a SGML/XML to HTML Java and JavaScript conversion script which will be integrated into the presently existing U.S. Navy technical document management system to produce HTML files. SGML components may be maintained and/or authored utilizing the Arbortext Adept series,

stored in the Chrystal Astoria component level Document Manager, and then exported to the file system as XML. The SGML/XML to HTML Java and Java Scrip conversion script in accord with the present invention convert the XML documents into the HTML format in accord with examples shown in FIG. 1 and FIG. 2 which disclose how SGML attributes and tags are mapped to an HTML equivalent.

[0033] In a preferred embodiment, the presently improved system utilizes open architecture, non-proprietary systems and method (Java and JavaScript) which enables entire system to be migrated from developer to developer without having to purchase proprietary systems, re-train developers, or re-author the source content to fit the new developers' inherent system.

[0034] The SGML/XML to HTML Component Conversion Method combines open architecture, non-proprietary Commercial Off The Shelf (COTS) authoring and storage products with developmental items. Some aspects of a preferred embodiment the system include an ODBC compliant object oriented database and the ultimate format of the baseline technical documentation data in SGML and XML. The display is provided using one of the common, freely distributable browsers (i.e., HTML output). In the resulting conversion, the functionality of Interactive Electronic Technical Manuals (bookmarks, annotation, full text search) is supported.

Various features include a frames based display, classification banners, and an expandable Table of Contents (TOC) with word wrap (see FIG. 3 and FIG. 4).

[0035] In one embodiment, the method comprised of an Arbortext Adept Series is used to edit and publish the SMGL/XML components but any SGML editor will suffice. A Xerox Document Management groups Chrystal SGML component database management product may be used but any ODBC compliant object oriented database capable of handling SGML components will suffice. Various Java and JavaScript authoring tools may be utilized. Baseline commercial utilities may include a parsing engine and XML library.

[0036] The conversion process initiates by defining an overall "Class" variable that determines the conversion class, such as for any SGML DTD. The (DTD) Document Type Definitions provide definitions for each tag, entity, and attribute for the conversion class. In one case, the class may define the mappings to HTML as methods that are called from an external XML parsing engine when the document starts, ends, and then every time any element, comment or processing instructions are encountered in the source document or fragment. Possible examples of such definitions for a class of SGML tags are shown in FIG. 1 and FIG. 2. As indicated in FIG. 1 and FIG. 2, the Java/JavaScript executable converts each method (tag, entity, and attribute) to HTML in accordance with the HTML 3.0. DTD. In one embodiment,

the SGML/XML tags, entities and attributes are formatted in accordance with the NAVSEA02 DTD. The output files are then combined with the graphic images, which may be separate .pdf files, in the appropriate folder structure to yield the final web viewable electronic document.

[0037] The conversion application checks attribute values as indicated in FIG. 2 of appropriate elements during the conversion process. Each chapter, index, appendix, and glossary is divided into a separate HTML file during the conversion process. The conversion process creates four types of frames: content frame 12, global navigation frame 13, local navigation frame 16, and classification frame 22.

[0038] Content frame 12 is preferably located at the center right of browser display 10. The content frame displays the content of a specified book, i.e., a converted HTML document. Vertical scroll bar 14 may be utilized to scroll up and down for navigating this frame.

[0039] Local navigation frame 16 is located to the left of the content frame. Local navigation frame 16 displays Table of Content (TOC) 18 of the section currently displayed in content frame 12. As the user moves within content frame 12, the appropriate TOC displays in local navigation frame 16. Vertical scroll bar 20 may be utilized to scroll up and down to navigate this frame. Preferably, TOC 18 is created by using collapsible

tree view which is accomplished using either JavaScript or a Java applet.

[0040] Global navigation frame 13 is shown at the top of the browser display. The currently open file is preferably displayed prominently in the frame 13. This frame 13 provides links such as those shown for SEARCH, VIEW BOOKMARKS, VIEW NOTES, and HELP that are available and applicable to all sections, not limited to the section currently open in the content frame.

[0041] Classification frame 22 is located at the top and bottom of the browser display 10. Classification frame 22 displays the highest classification of the content current HTML document being displayed in content frame. All possible classifications in descending order are: Secret (S), Confidential (C), and Unclassified (U). If desired, classification frame or another frame may be utilized to display on the highest classification of what is currently being displayed on the screen in classification frame.

[0042] Graphics 24 can appear inline within content frame 12 or as a separate pop-up window of the application spawning the graphic, e.g., Adobe acrobat, Internet Explorer, and the like, and may contain a separate classification banner which is generated from the security attribute as indicated in FIG. 2.

[0043] Bookmarks, annotations, and local and global search functions appear in separate Internet Explorer pop-up windows.

Bookmarks may be generated by selecting the bookmark icon 26 which appears in the HTML after each document section. The mechanisms for generating bookmark icon 26 comprise providing a unique SGML tag after each "para0" and "subpara1" which the filter migrates to a hyperlinked icon. When the user wants to create a bookmark, he simply selects bookmark icon 26 and a bookmark is generated. When the user wants to retrieve a bookmark, he selects the "bookmark" hyperlink in the local or global navigation or table of contents frame 16 which generates a new browser instance pop-up window that displays all the bookmarks generated during the user defined session that the user can then select to display the bookmarked content in the same window.

[0044] Annotations are generated by selecting annotation icon 28 which appears in the HTML after each document section. The mechanisms for generating annotation icon 28 comprise utilizing a unique SGML tag after each "para0" and "subpara1" which the filter migrates to a hyperlinked icon. When the user wants to create an annotation, he simply selects annotation icon 28 and an annotation is generated. When the user wants to retrieve an annotation, he selects the "annotation" hyperlink in the local or global navigation or table of contents frame 16 which generates a new browser instance pop-up window that displays all the annotations generated during the user defined session and also

the content of the annotation. The user can then select the text that the note is anchored to which displays the full context against which the note is generated in the same pop-up window.

[0045] Search (both local and global) is initiated by selecting search hyperlink 30 or 32 in the respective local or global navigation frame, which generates a pop-up window that prompts the user to enter the string to be searched. The search engine then combs either a local (current document) or global (all documents in the collection) text (.txt) files which contains ASCII text characters for all content, and then displays the results (with hyperlinked text) in a new browser instance pop-up window.

[0046] It is to be noted that each of the pop-up windows (for bookmarks, annotations, and search) preferably contains a classification banner frame (above and below the destination data) generated from the "Security" attribute nested in the SGML baseline.

[0047] The current method preferably translates "Warning", "Caution", and "Note" SGML tags to HTML specific formatted text complete with the appropriate Icons. The present invention preferably provides Hyperlinks (HTML) for all SGML tags and attributes ("xref", "xerfid", "Glossary", and "term") for intra and inter document links, Glossary, and term and definition links.

[0048] A combined Java and JavaScript software program in accord with the present invention translates SGML/XML tags, entities, and attributes from the source document to HTML standard (WWW HTML 3.0) tag sets that are meaningful and viewable in at least the MSIE 4.01 Service Release (SR) 1 web browser as indicated in FIG. 1 and FIG. 2.

[0049] In FIG. 1, Tag map 34 shows SGML tags 35 which are converted into HTML viewable tags 36 with a Java filter. In FIG. 2, SGML/XML attribute map shows how the elements are to be mapped into HTML. For instance, body elements 38, which may comprise paragraph tags (para0, subpara1), figures, tables, etc., are mapped into a Java/JavaScript produced table of contents as indicated at 40. As noted above, the conversion application checks attribute values, such as values 42 of appropriate elements during the conversion process. FIG. 3 shows the tags 44 and attributes identified in FIG. 1 and FIG. 2 as applied to the new HTML browser viewer. Each chapter, index, appendix, and glossary is divided into a separate HTML file during the conversion process. Thus, in a preferred embodiment as noted above, the conversion process creates four types of frames: content frame 12, global navigation frame 13, local navigation frame 16, and classification frame 22.

[0050] The electronic document users of the converted HTML documents allows users to view their documents in a uniquely

presented frame based HTML environment 10 with full functioning reference links including Glossary, terms, and definition references. As well, the user may search all text in a local or global mode. It is also possible for the user to link to/from the results and to create annotations and bookmarks in the HTML document. The process/product also auto-generates TOC (table of contents) structure data preferably using a combined Java/JavaScript runtime executable.

[0051] Thus, the present invention provides uniquely archived features of bookmarks, annotations, and full text searches from within web browser interface 10 but maintained in the SGML/XML base line. The baseline data is formatted in a hierarchical, object-oriented database, which allows for the reuse of common data. This is a great cost benefit/savings which enables developers to change source content one time in the baseline even though it may appear in numerous locations. The content may be used with other training sources. The source content of SGML components, which can be reused as source, and the final output of HTML files, can be easily re-used with other training systems as stand-alone chunks as long as a compatible web browser is available.

[0052] Alternative embodiments may comprise using extensible Style Language (XSL) or cascading style sheets along with Microsoft's MSXML DLL. Another approach employs a customized

version of the conversion script authored in a programming language other than Java, e.g., C++, VB, and the like.

[0053] There are many documents authored in SGML in other industries (e.g., law, publishing) for which there are no free viewers for SGML. However, the present method formats the data to be viewable in the Microsoft Internet Explorer 4.0 or greater browser configuration which for all intents and purposes is free complete with additional functionality of searching all content in standalone (not web server) mode and allows for bookmarking and annotating all data in the collections (even to a specific location within an HTML file).

[0054] In summary, the present invention provides a mapping for Military Document Type Definition (DTD) specific components, which address unique attributes and elements not provided by commercial DTDs. Some (but not all) of the elements and attributes incumbent in the NAVSEA02 DTD include content classification (unclassified - FOUO, Confidential, Secret, Top Secret). In this case, this attribute is parsed and represented by a dynamic frameset component that maps to the highest classification of the active HTML, graphic, or multimedia file or document.

[0055] The present invention provides mapping chapter, section, para, subpara, Figure, foldout, table, multimedia components to multi-dimensional (expandable/collapsible) Table of Contents (TOC) 18 integrated frameset.

[0056] The present invention provides a bookmark feature that dynamically anchors a return feature to any translated SGML/XML component in the HTML. The process requires that a unique SGML component and attribute be defined in the source data that is mapped to icon classes, and executable application code (JavaScript or Java for example) that, when enabled, execute the bookmark write and retrieve features, via separate code execution when activated. Additionally, the method integrates the executable code that when activated, spawns the query interface which allows the use to interactively query the subject domain for populated bookmarks.

[0057] An annotation feature dynamically anchors a return and user defined, interactive note feature to any translated SGML/XML component in the HTML. The process requires that a unique SGML component and attribute be defined in the source data that is mapped to icon classes, and executable application code (JavaScript or Java for example) that, when enabled, executes the annotation and note append, write, and retrieve features, via separate code execution. Additionally, the method integrates the executable code that when activated, spawns the query interface

which allows the use to interactively query the subject domain for populated annotations.

[0058] The present invention provides for local and global full text search. The SGML plain text file is parsed and the structure elements are stripped out to generate a plain text file. The local feature interrogates a single document (SGML source file) domain and the global search interrogates a multi-volume (configurable number of SGML source files) domain, parses, and collates into separate document volumes and displays associated meta data components housed in the SGML source. Additionally, the method integrates the executable code that when activated, spawns the query interface which allows the user to interactively query the subject domain for matches and which then dynamically escalates to the results interface.

[0059] The invention provides for dynamic handling of reverse links from the graphics to textural information. The method converts and dynamically manages location and specific vector location hyperlinks. The method dynamically manages the HTML representation by combining style sheets and application code to render a configurable multi-frame format. The method dynamically configures the translation from the source file information which includes dynamically managing the structure components and the scope of the items to consider in an individual or automated sequence of application runs.

[0060] Thus, the present invention provides for conversion of an SGML formatted document, with associated graphic, multimedia to configurable HTML stances. The parser allows for multi-layered HTML documents comprised of dynamically sizing framesets which combines transformed SGML and external components such as graphics, multimedia, and executable mini-applications that enable external (to the SGML document domain) referencing any add-on interactive features such as local and global search, bookmarks, and annotations. The parser translates all SGML structure specific attributes to unique HTML elements.

[0061] It will be understood that many additional changes in the details, materials, steps and arrangement of parts, which have been herein described and illustrated in order to explain the nature of the invention, may be made by those skilled in the art within the principle and scope of the invention as expressed in the appended claims.

SGML/XML TO HTML CONVERSION SYSTEM  
AND METHOD FOR FRAME-BASED VIEWER

ABSTRACT OF THE DISCLOSURE

A system and method provides conversion of SGML or XML files to HTML files by component conversion techniques. In one embodiment, the system addresses the conversion of an SGML formatted document, with associated graphic and multimedia for viewing in a plurality of frames produced in a web browser, such as with Java or JavaScript. A parser allows for multi-layered HTML documents comprised of dynamically sizing framesets which combines transformed SGML and external components such as graphics multimedia with executable mini-applications that enable external (to the SGML document domain) referencing of add-on interactive features such as local and global search, bookmarks, and annotation. In a preferred embodiment, a Java or JavaScript executable produces a collapsible/expandable table of contents in one of the frames.

SGML TAG/TAG GROUP		HTML EQUIVALENT	
START	END	START	END
<p>35</p> <p>&lt;pubno&gt;&lt;docno&gt; &lt;doctype&gt; &lt;prtitle&gt;&lt;nomen&gt; &lt;mfr&gt; &lt;pubdate&gt; &lt;?Pub newline?&gt; &lt;emphasis&gt; &lt;emphasis emph=b&gt; &lt;para&gt; &lt;chapter&gt;&lt;title&gt; &lt;forward&gt;&lt;title&gt; &lt;preface&gt;&lt;title&gt; &lt;appendix&gt;&lt;title&gt; &lt;seqlist&gt; &lt;randllist&gt; &lt;item&gt; text string '(U)' text string '(C)' text string '(S)'</p>	<p>&lt;/pubno&gt;&lt;/docno&gt; &lt;/doctype&gt; &lt;/prtitle&gt;&lt;/nomen&gt; &lt;/mfr&gt; &lt;/pubdate&gt; - &lt;/emphasis&gt; &lt;/emphasis&gt; &lt;/para&gt; &lt;/title&gt; &lt;/title&gt; &lt;/title&gt; &lt;/title&gt; &lt;/seqlist&gt; &lt;/randllist&gt; &lt;/item&gt;</p>	<p>&lt;H3 ALIGN="CENTER"&gt; &lt;H3 ALIGN="CENTER"&gt; &lt;H2 ALIGN="CENTER"&gt; &lt;H6 ALIGN="CENTER"&gt; &lt;P ALIGN="RIGHT"&gt; &lt;BR&gt; &lt;EM&gt; &lt;STRONG&gt; &lt;P&gt; &lt;H2 ALIGN="CENTER"&gt; &lt;H3 ALIGN="CENTER"&gt; &lt;H2 ALIGN="CENTER"&gt; &lt;H4 ALIGN="CENTER"&gt; &lt;OL&gt; &lt;UL&gt; &lt;LI&gt; generate hidden 'UNCLASSIFIED' tag generate hidden 'CONFIDENTIAL' tag generate hidden 'SECRET' tag</p>	<p>&lt;/H3&gt; &lt;/H3&gt; &lt;/H2&gt; &lt;/H6&gt; &lt;/P&gt; - &lt;/EM&gt; &lt;/STRONG&gt; &lt;/P&gt; &lt;/H2&gt; &lt;/H3&gt; &lt;/H2&gt; &lt;/H4&gt; &lt;/OL&gt; &lt;/UL&gt; &lt;/LI&gt;</p> <p>36</p>
TAGS REPLACED WITH ATTRIBUTES			
TAGS REPLACED WITH ATTRIBUTES			

FIG. 1

ATTRIBUTE	SGML TAG	VALUE	MAPPING/DESCRIPTION
TOCENTRY	BODY ELEMENTS	1	ELEMENT TO BE INCLUDED IN TABLE OF CONTENTS.
ID	BODY ELEMENTS		STRUCTURE NOMENCLATURE - UNIQUE IDENTIFIER (NOT PRINTED)
LABEL	BODY ELEMENTS	TEXT STRING	TITLE MAPPED TO HTML <H1> THROUGH <H8> SEE THE TAG MAPPING IN FIG 1
EMPH	EMPHASIS	B, BO, J	
XREFID	XREF, EXTREF	TEXT STRING	EQUALS THE 'ID' OF THE DESTINATION ELEMENT
BOARDND	GRAPHIC	TEXT STRING	ENTITY NAME OF GRAPHIC
GRAPHSTY	GRAPHIC	INLINE/OUTLINE	USED TO DEFINE INLINE GRAPHIC OR POP-UP WINDOW
ALIGN	ENTRY	LEFT, CENTER, RIGHT	TEXT ATTRIBUTES IN TABLES
SECURITY	BODY ELEMENT	U=UNCLASSIFIED, C=CONFIDENTIAL, S=SECRET	CLASSIFICATION APPEARS IN HEADER AND FOOTER FRAMES

40

38

42

FIG. 2

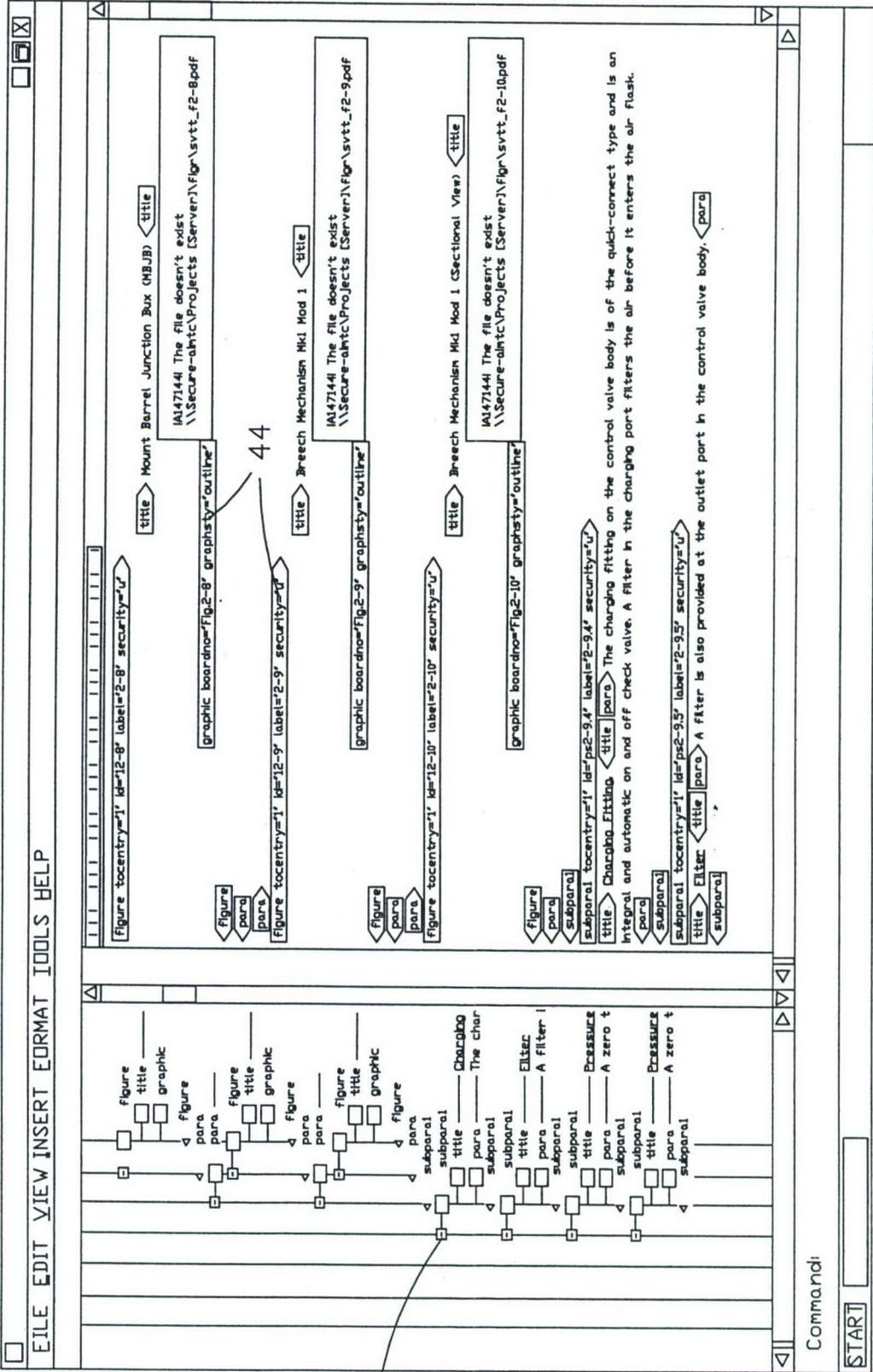


FIG. 3

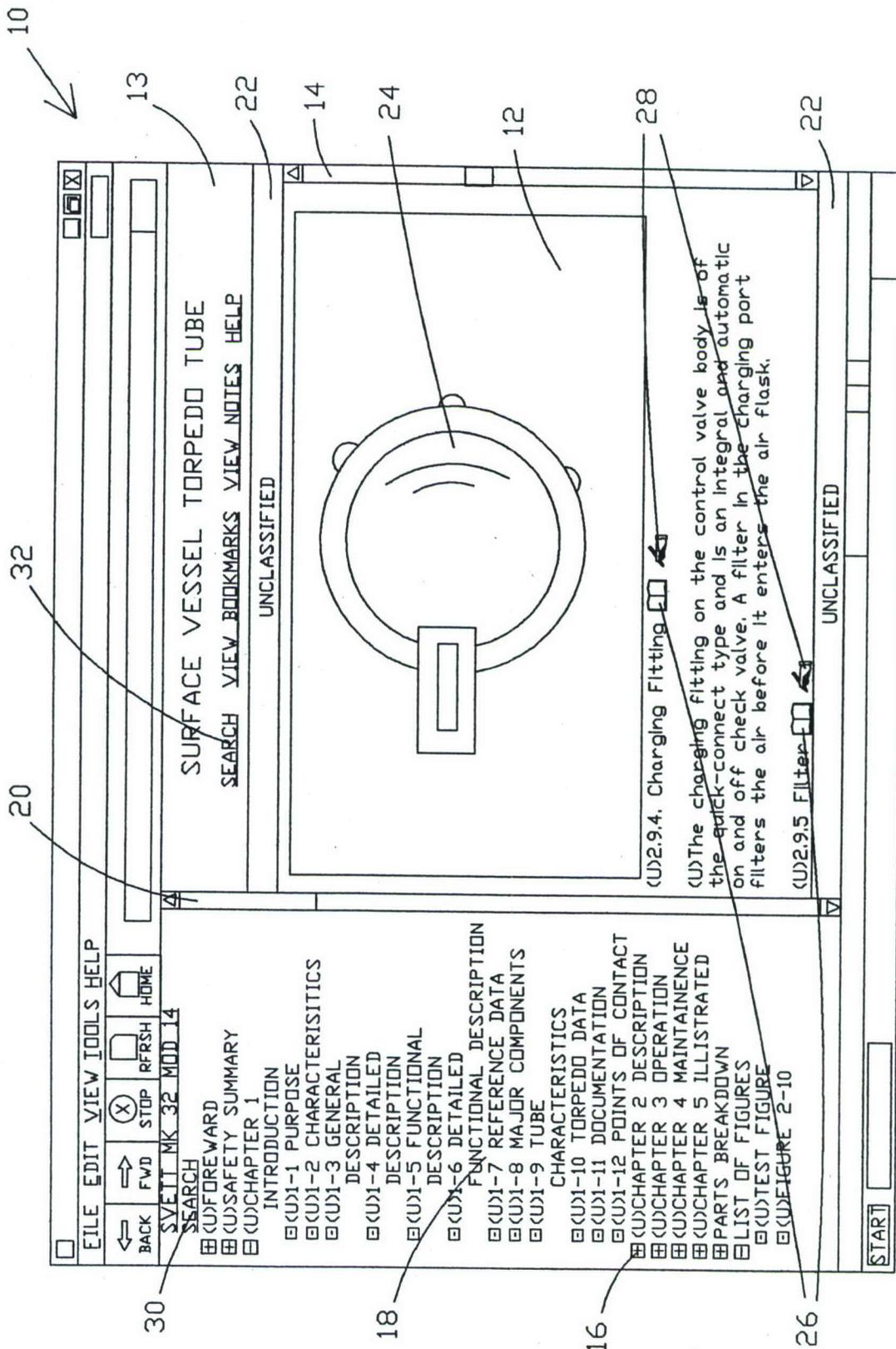


FIG. 4