Abstract

NYNEX Information Resources Company has assisted in developing a system that integrates gridding guidelines and expertise with conventional programs in order to solve an increasing business problem. DX, Directory Expert, is a Lisp based system that merges gridding specifications with page layout and graphic programs to create the Yellow Pages. This system employs a custom built inference engine that not only drives the knowledge components but also applies domain specific knowledge to optimize the page layout process. As pages are generated the system tests each configuration based upon NYNEX Gridding Specifications. The resulting pages are automatically evaluated so that the gridders can concentrate on correcting the difficult exceptions to the gridding specifications.

1 Introduction

The NYNEX Yellow Pages provide a valuable service to millions of customers each day. NYNEX Information Resources Co., NIRC, is a respected leader in the Yellow Pages industry. Although much of NIRC's Yellow Pages are compiled using advanced computer technology, one of the key steps currently used involves a manual process - the placement of the Yellow Page ads and listings onto individual pages, referred to as gridding. Gridding the NIRC Yellow Pages requires a knowledge of NYNEX Gridding Specifications while optimally positioning all items to fit on a page. Due to the high transaction volume and the complex specifications, gridders can benefit from a mechanism to assist them with the layout of the Yellow Pages. Directory Expert, DX, is such a mechanism. It captures, preserves and applies the expertise involved in laying out the Yellow Page Directories.

DX will replace this manual process with an automated pagination system which will not only handle all of the layout tasks but will also encompass other associated functions which are presently done by additional personnel. DX is the NYNEX pagination system of the future.

2 Background

The DX system was originally contracted out for full implementation. NYNEX provided the data necessary to build the system and worked with the contractor to develop a set of acceptance criteria. Unfortunately, the contractor defaulted on the project, leaving NIRC with the source code of a partially developed system with great potential. NIRC contracted AGS Information Services (AGS) to determine the time and effort it would take to complete DX. A DX Development Team comprised of NIRC and AGS was assembled to develop and deploy a production-ready DX system.

2.2 Existing Process

The proofread listings and ad information are used to generate a galley data file. This file is printed out, corrected manually, and given to gridders. The gridders manually compare this to an ad registrar, a prioritized list of ads. Any discrepancies become late queries, which are manually checked for validity and gridded late in the process. As the pages are gridded, the Page Intelligence (PI) division takes the grids and enters the data coordinates into the computer. When the grider receives a response to the query, the page will be re-grid and re-entered at PI, or let go depending on the lateness in the day and daily quota.

Once all pages are gridded and PI-ed, the pagination system produces Finals. The copied finals are forwarded to the Page Make-up area along with the
grid and the Finals, where the additional changes are inputted. The responses to the query can come back at any time throughout the Page development process. The page that contains a query is located, and a deletion is attempted with the minimum impact. Most of the time the deletion effects only one page; thus, it is handled by Page Make-up division. However, gridders assist when an ad needs to be deleted that will cause ripple. This is all done manually, by cut and paste. The PI and pagination are not involved in the process, since this can cause loss of time in the work flow.

3 DX System Process

DX is an independent pagination tool which is based upon custom corporate layout specifications. A discrete pagination knowledge base, much of which has been incorporated into the DX inference engine, defines the basics of the generic pagination process. The advantage to building this type of system is that use by multiple users may generate different results, based on the individual corporate specifications. A user can also develop different types of books (e.g. 2 and 4 column books) by subdividing the knowledge into smaller units and using that which is similar at a higher level and adding additional knowledge at a lower level. The disadvantage with this type of system is that it is harder to develop, since it is more generic. But the tradeoff is in favor of the user(s) and that is why this system is so important.

The NIRC guidelines form the basis of the knowledge base. During the manual gridding process these guidelines are subject to interpretation by the individual gridders. DX will produce more consistent books in a more timely fashion.

This is accomplished by applying a set of page layout rules elicited from the expert gridders through a specialized interviewing and prototyping process called "knowledge engineering." These rules are the practical interpretation of the Gridding Specifications applied to actual data. When the DX system is able to replicate reliably the performance of expert gridders in an acceptably high percentage of cases, it will replace the current manually intensive process with an automated process that only needs to be "fine tuned" by an expert grider in those few cases that it lacks the expertise to handle properly.

3.1 DX System Flow

DX reads the galley file of proofread listings and ad information and determines how to arrange the ads and listings on pages and spreads (sets of two facing pages) in accordance with the policies and procedures elicited from expert gridders through knowledge acquisition and according to the NYNEX Gridding Specifications Manual. Since most pages cannot be completely filled in with just ads and listings, it is necessary to employ filler ads and/or listings to fill up what would otherwise be blank space. The output of the system, therefore, consists of ad, listing and filler placement information combined with the original galley file information. This output can then be used to produce the final pages constituting a Yellow Pages Book.

3.2 DX System Architecture

DX is a Lisp-based system which runs on TI Explorers and employs a robust user interface for page layout. DX resembles a traditional expert system in that it uses an inference engine, knowledge base, and supporting procedural code. DX also uses multiple knowledge bases and associated driver routines to serve both the input and output capabilities of the system and to link DX to the other parts of the computer-managed process. The DX knowledge base is arranged as a hierarchy of hookpoints (rule groups linked by functionality) which are supported by procedural rule primitives.

The user interface comprises a graphical interface to allow the users to select a galley file (the data file consisting of the materials that make up the Yellow Pages) and to run a book. After the initial run the user can use a manual editor facility to correct any pages that are not deemed acceptable. A bad page identification mechanism has been developed which focuses the gridders attention to only those pages that do not meet the specifications. This prevents the user from having to review all the pages to identify and correct an acceptable page.

The inference engine serves two purposes: to apply rules and to constrain page layouts. That is, aside from validating and firing rules, the inference engine has built in domain knowledge to allow only certain page layouts. Though this speeds up execution time, by not generating unnecessary illegal paths, and by reducing the number of rules needed to select legal pages, it does require additional knowledge for its modification. The
Rule firing mechanism is comprised of hookpoints, an enabling pattern to segregate rules into rule groups. This allows only those rules that are relevant during certain situations to be considered at any point in time. Rules are divided by logical gridding steps: spread apportionment, division apportionment, ad placement, listing placement, and filler selection.

The inference engine uses a Best First Search mechanism which includes generate and test to pick page candidates using the page scoring mechanism. Once the best candidate is chosen, further refinements are spawned from this page.

Page Scoring is based on a weighted average, using the rule weight and the rule value. This score is calculated after each rule group has fired and the page candidate with the highest score is chosen to proceed with the layout. This generates a tree with several parent and children nodes. The final leaves of the tree are either legal or failed alternative layouts. The legal layout with the highest final score is the chosen page.

A blackboard architecture is used extensively by the inference engine. Messages are left by rules or functions in the blackboard and when necessary are picked up by the look-up functions that periodically check the blackboard. As a result the rule groups can be instantiated over and over until a certain condition is met.

Rules are converted into a compiled database and are dynamically generated into rule groups which decreases execution time. Rules are applied and fired in several modes (normal mode, relax mode, quick and dirty mode). In normal mode the rules are as specified in the specifications manual. Relax rules allow for exceptions that are in the subjective areas of the specification. Quick and dirty mode requires manual editing for specification overrides. These modes of operation occur consecutively until a legal spread is laid out.

Whenever the contents of a finished page needs to be changed for any reason (e.g. adding or removing an ad or listing), a condition called ripple may occur, whereby the effected page and the pages immediately surrounding the change are reconfigured.

To accommodate late corrections and to minimize ripple a late queue facility, used for galley additions, deletions and corrections, has been developed to merge late items with the input galley file. This facility has the effect of minimizing the need for rippling late changes through previously completed good pages.

After running DX and making modifications with the manual editor the gridders generates an output galley file. This file contains the coordinates and control information for all items in the book.

3.3 DX System Integration

DX is designed to interface with existing computer and graphic systems used in the Yellow Page creation process. DX input consists of galley data and late galley data corrections, additions, and deletions. These files are merged during the read operation by the sequential galley number of each item. DX generates output, in the form of data known as galley files, for the GRAFX system and then transfers its final results to photocomposition. DX provides, as an option, an interactive display for the user to view the pagination process during layout. Most often DX will run pages during off hours, without user monitoring. The resulting pages may then be reviewed and modified by the user as deemed necessary. Using DX the gridders will review the generated pages; using a system option DX will advise the gridders which pages may not be acceptable, based upon predetermined constraints and the use of a pagination self evaluation mechanism - BPID. With DX the user no longer uses the manual "cut and paste" method, but does this using the interactive display, a mouse and the DX manual editing capability.

4 Verification and Validation

Some obstacles concerning page validation and expertise had to be overcome. To satisfy best customer requirements specifications continually are upgraded; however, the specifications are often subject to interpretation. This caused difficulty with knowledge formalization. An agreement was made to follow the expert's decision-making process since the practice supersedes the theory. By constantly gridding pages the rules evolve. The expert supervised his personnel to produce daily quota of pages, therefore he was not always available as needed. However, the expert's enthusiasm and willingness to cooperate compensate for the limited access time.

Initially there were several sources of expertise, however none of them was an "expert" in the true sense of the definition. These highly qualified individuals where deemed as "advanced users" and their expertise was incomplete and conflicting. Additionally the specifications manual was not specific enough and
subject to interpretation. This caused difficulties in understanding the rules initially during the acquisition; however, the problem was reduced significantly during the rule validation phase. The above difficulty was lessened once NYNEX gave the ownership of the system to one individual. Although, since the specification is a growing document, a new problem arose that needed proper management and timely upgrades. The conflicts in rule validations would be apparent when the newly acquired rule would become part of the rule base prior to being updated in the manual. These conditions made rule verification harder to achieve.

To ease the transition for gridders to use DX two types of parallel tests have occurred: development and production parallel tests. Development parallel was used to check the user interface, any potential "bugs" in the system, and to familiarize the gridders with using DX. The production parallel test was used to compare the results of manually gridded pages with the results of DX. All features of DX were used such as the add/delete facility (whereby a late item can be inserted into the book after the pages have been run) and the manual editor.

DX is currently in extensive production parallel test. Expert gridders have been trained on the system and use DX while their staff are manually gridding books. The current goal is to have DX grid 150 pages per day: the same goal given to the current gridding teams.

To verify and validate DX further, several tools were developed. These generic facilities include dynamic page traces, no rule mode analysis, page tree traversal, rule failure logging, chosen page analysis, and bad page identification.

5 Benefits

DX has numerous benefits: it enables gridders to concentrate only on those situations deemed difficult; produces a more consistent book layout amongst all gridders; reduces the size of a book by applying filler as needed; speeds up the gridding process time especially when more than one book is gridded simultaneously; takes less time to correct pages by merging late items with the initial input file.

6 Conclusion

The generic design of the DX inference engine lends itself naturally to implementation of new gridding policies. This flexibility allows changes in the nature of book layouts by simply changing expert rules resident in the knowledge base. This layout modification, combined with the ability to easily accommodate late items, gives NYNEX a competitive advantage in responding to the demands of the various regional markets.

DX provides consistency of application to the various NYNEX policies for gridding the different type of Yellow Page Books. Consistency of production will go a long way toward ensuring that NYNEX customers continue to enjoy high levels of quality service.

Acknowledgements

The authors greatly acknowledge the invaluable assistance of the following people: Thomas Cain, Milton Eleftheriou, Ray Hrbek, Eileen Brannick, Lanny Grim, and Kathy Van Bergen.