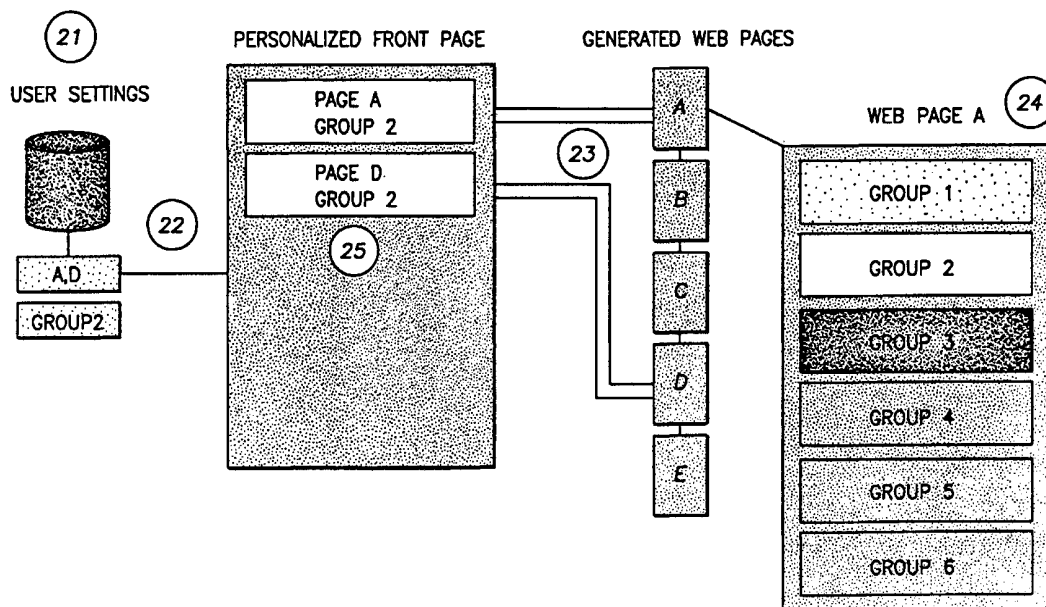




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>7</sup> :</b>  <b>G06F 17/30</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 00/62206</b>  <b>(43) International Publication Date:</b> 19 October 2000 (19.10.00)
<b>(21) International Application Number:</b> PCT/US00/09974 <b>(22) International Filing Date:</b> 13 April 2000 (13.04.00)  <b>(30) Priority Data:</b> 60/129,106      13 April 1999 (13.04.99)      US  <b>(71) Applicant:</b> CONJOIN, INC. [US/US]; Suite 355, 20 Mall Road, Burlington, MA 01803 (US).  <b>(72) Inventors:</b> DIMARE, Joseph; 197 High Street #12, Andover, MA 01810 (US). HEATH, Barbara; 28 Village View Road, Westford, MA 01886 (US). D'ARBELOFF, Nicholas; 345 Cross Street, Belmont, MA 02178 (US).  <b>(74) Agent:</b> NUGENT, Elizabeth, E.; Choate, Hall & Stewart, Exchange Place, 53 State Street, Boston, MA 02109 (US).		<b>(81) Designated States:</b> AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>

**(54) Title:** GROUP TARGETED CONTENT PERSONALIZATION**(57) Abstract**

Group Targeted Content Personalization (GTCP) is a multi-dimensional categorization of content taxonomy and user-affiliation which generates personalized web home pages for individuals. The web pages are personalized both by the audience type of a user and by the personal preferences of the user. A content database may include measurements of content quality, which can be used to sort displayed content for the user.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## Group Targeted Content Personalization

The present application claims priority to U.S. Provisional Application No. 60/129,106, entitled "Group Targeted Content Personalization," filed April 13, 1999.

5

### Field of the Invention

The invention pertains to a system for generating personalized listings of relevant content to users.

10

### Background of the Invention

As Internet- and intranet-based computerized systems for information delivery become more common, personalization of systems has become both more popular and more difficult. As the number of users of a system grows, the difficulty of storing individual preference information and of dynamically responding to a request for personalized information increases dramatically. At the same time, consumers have increased demand for personalization features. A large number of providers of "personal" content pages such as "My Yahoo!"<sup>®</sup> and "my.CNN.com"<sup>™</sup> have emerged, but their services are typically either only marginally personalized, or are unacceptably slow. A need thus exists for an improved method of information delivery, which can supply targeted content which is relevant to the needs of a particular user. In addition, most available "personal" content pages cannot address document security. A need exists for a personalized information delivery system that can discriminate by user type and deliver only information that the user is authorized to access.

25

### Summary of the Invention

Group Targeted Content Personalization (GTCP) is a multi-dimensional categorization of content taxonomy and user-affiliation which generates personalized web home pages for individuals. Some web sites have classified content by taxonomy and other sites have associated individuals with groups. GTCP transcends both by combining and cross-referencing these categorizations into a multi-dimensional view.

30

Secondly, GTCP is implemented using a new methodology for applying pre-generated or static components to present an essentially dynamic collection of information. Many web sites generate static information in order to speed display of such information. Dynamically generated information (typically as a result of database queries) is more suited to customization but does not scale well with a large number of users. GTCP combines both methods in a new way by generating static information based on both the content taxonomy and the group classification and then combines this multi-dimensional static information dynamically to the user.

Additional value to the user is provided by the ordering of the display based on prior feedback from other users. This content lifecycle management information prioritizes the GTCP display, allowing more highly valued content to be displayed first, further enhancing the targeting aspect of personalization.

Finally, the combination of content taxonomy and group affiliation in GTCP lends itself ideally to the presentation of information to the user in several ways. Information may be access controlled by group or targeted by content to group, or combination of both.

In one aspect, the invention comprises a computer-implemented method for generating a personalized information listing. The listing includes identifiers for content items, which may be, for example, URLs, file location paths, directories or subdirectories, or even catalogue information for nondigital information such as library references for books. The method includes generating a set of static content pages associated with each of a number of content areas, and then subdividing the static pages according to access privileges for various audiences. When a user requests a personalized listing, the pages corresponding to that user's interest areas are selected, and only those portions of the pages corresponding to the user's access classification are returned. In preferred embodiments, the identifiers are sorted according to content value. In particular, content items may have one or more designated target audiences, and the subdivisions of the static pages may be sorted to place items of particular interest to the associated audience at the top of the page. In addition, page contents may be sorted by value ratings and/or usage ratings in order to concentrate valuable items at the top of the list.

### **Brief Description of the Drawing**

The invention is described with reference to the several figures of the drawing, in which,

5       **Figure 1** is a flow chart showing the generation of static pages as a function of content and user categorization;

**Figure 2** is a flow chart showing the generation of a personalized page based on user classification and user preferences;

**Figure 3** illustrates the speed improvements which may be made by reducing the load on the server using the methods of the invention;

10       **Figure 4** shows a customized screen presented to a particular user by the methods of the invention; and

**Figure 5** shows a preference screen where a user may update his or her preference data.

15       **Detailed Description of Certain Preferred Embodiments**

      Group Targeted Content Personalization (GTCP) is a multi-dimensional categorization of content taxonomy and user-affiliation which generates personalized web home pages for individuals. Some web sites have classified content by taxonomy and other sites have associated individuals with groups. GTCP transcends both by  
20       combining and cross-referencing these categorizations into a multi-dimensional view.

      Secondly, GTCP is implemented using a new methodology for applying pre-generated or static components to present an essentially dynamic collection of information. Many web sites generate static information in order to speed display of such information. Dynamically generated information (typically as a result of  
25       database queries) is more suited to customization but does not scale well with a large number of users. GTCP combines both methods in a new way by generating static information based on both the content taxonomy and the group classification and then combines this multi-dimensional static information dynamically to the user.

30       Additional value to the user can be provided by the ordering of the display based on prior feedback from other users. This content lifecycle management information prioritizes the GTCP display, allowing more highly valued content to be displayed first, further enhancing the targeting aspect of personalization. A

compatible system for determining value of content is more fully described in U.S. Provisional Application 60/129,104, filed on April 13, 1999, and in U.S. Patent Application "Content Lifecycle Management," filed on even date herewith.

5 Finally, the combination of content taxonomy and group affiliation in GTCP lends itself ideally to the presentation of information to the user in several ways. Information may be access controlled by group, targeted by content to group, or a combination of both.

GTCP delivers targeted and general content to a user in a personalized view based on specific information that the user has supplied about their interests.

10 Personalization is viewed as a function both of the categorization of the content and of the users viewing the content. Four variables that may be used to control the ability to view the content are: (1) the user's audience type, which controls access to the content as well as targeting of the content (*e.g.*, a typical audience type might include members of a specific department in an organization, or members at a

15 particular management level); (2) the content's rating and usage, which relates the value and benefit of the content (and may be dependent on the audience type); (3) the date range of viewable content and the content expiration; and (4) the user's interest in the content, as determined by user-specific settings.

One embodiment of the invention uses a two-tiered business taxonomy

20 classification of topics and sub-topics (*e.g.*, stored in a relational database) to generate static web pages. The total number of pages generated is based on the number of entries in the subtopics table. The generation of the pages is a scheduled task that can be run automatically at a site-defined interval. Each page contains groupings of the viewable content from the library table by audience type derived from the audience

25 table. These are **classification-encapsulated static pages (CESP)**. Within each of the groupings, content that is specifically targeted to the user's audience type is displayed first because of the importance to the user. General content which is viewable by the user, but not specifically targeted to the user, follows the targeted information. And finally, the content within the targeted and general sub-groupings is

30 further stratified using the contents' ratings and usage from the ratings table and usage table. The benefit to the user is immediate as content that is both targeted to the user and of most value is positioned at the top of the returned list.

Websites consist of static information and dynamic database-driven information. Static content refers to the display of non-changing text-based pages. Database-driven web content refers to a real-time process of querying data from a database given specific parameters to search for and then returning the matching records for display. Today, many sites use databases to query data and then generate static pages for display. Large commercial sites use this method in order to reduce the load of accessing the database by multiple users. However, in these cases, the entire dynamically-generated static page is shown to all users who have chosen this page. There is no customization of the content within the generated (or static) file for an individual or group.

The Group Targeted Content Personalization extends the static page which is generated from a database to include an encapsulated-classification structure that maps content to user-specific groups (or audiences). Each generated page contains all the required information for all users but is specifically customized based on the audience type of the user who is viewing the personal page.

In addition to the grouping of content by targeted and general, there may be an additional stratification of the content within each page presented, based on a ratings and usage model as described in commonly-owned U.S. Provisional Application No. 60/129,104, filed April 13, 1999. This combination of scoring and usage allows the GTCP to place content of higher value (higher scores and more viewings) at the top. This also creates a constantly changing display of information based on the employees' feedback of the content.

### *Architecture*

In one embodiment of the invention, Group Targeted Content Personalization is based on a set of database tables that are part of an intranet database. A goal is to perform as many tasks as possible in the database application and then to generate the classification-encapsulated static web pages.

The following data are preferably cataloged in the database tables:

- Content information – specifically name, file name, content type, publish and expiration dates, approval and archive status
- Pointers to each document located on disk or web address (URL) for a link

- Access permission for each document by audience
- Targeting for each document by audience
- Content ratings submitted by users
- Content usage of each document
- 5      • User personalization settings (see **Figure 5**).
- Business taxonomy – topics and subtopics

The following data are also preferably stored on durable media such as a hard disk:

- 10      • Content (which may include documents, files, executables, or other materials)
- GTCP Daemon (code)
- Classification-encapsulated static web pages
- Personal page template(s)
- 15      • User settings template(s)

### *Content Publishing*

Content may be published to the intranet using a publishing functionality within the system. During the publishing process, a file is uploaded and stored on the server and all file-specific and publisher-entered information is saved in the database tables and associated with the uploaded file. Data specific content such as web links (URLs) are stored in the database and the information is associated with the database listing. The user-entered information may include the classification of the document using the topic and subtopic parameters, access classification (who has access to view the document), target classification (who would most benefit from this content), and expiration date.

### *GTCP Daemon*

The GTCP Daemon is the executable code which generates the classification-encapsulated static web pages. The daemon is a scheduled task which may be run on a periodic basis (defined by the system administrator) to generate a revised set of files

with updated content. The GTCP daemon queries the various tables in the intranet and generates the static pages. Specifically, the daemon will:

- Query the **subtopics table** for the total number of records in the table and loop through the records while performing the following:
    - 5                   • Capture the subtopic record id and name
    - Query the **audience table** for the total number of audience types in the table and loop through the records while performing the following:
      - Capture the audience ID and audience name
      - Query the **library table** for content that matches the current
        - 10               subtopic, audience ID, and target ID; and order the returned list using the ratings and usage tables
        - Return the ID, name, filename, abstract, date approved fields
        - Add additional code bounding the returned records that sets the audience-level security
        - 15               • Output returned information
    - Write-out the matching files from the library table for each audience type.
- Each CESP page generated from the subtopics table will have content grouped for each audience classification. By storing this content as a series of static pages according to the subtopics table, and then addressing audience-specific portions of these pages as described below, content which is personalized both by the interests and by the audience type of the user may be rapidly accessed and displayed. Typical savings in database load are shown in **Figure 3**.

#### *Classification-Encapsulated Static Web Pages*

- 25               Each CESP web page is a document containing all the viewable file references from the library table grouped by audience type and organized within each group by targeted content and general content. Each of these two groups may be further organized by valued content at the top and less valuable content at the bottom of each section. Each CESP web page will:
  - 30               • Check the user's server-side cookie file which tracks the identity of each user for the user's audience classification.

- CESP will then display the returned library records organized as cited above which match the audience type of the user.

### Example

5      *Generation of the Classification-Embedded Static Web Pages* (see **Figure 1**)

(1)      Query the Subtopics table 1

- Find all active records in the table and return the record set
- Set current record name and id as variables

10      (2)      Query the Library table 2

- Find all active records in the table and return the record set
- Set current record name and id as variables

(3)      Query the Library table 3

- 15      •      Find all matching records in library table that match:
- the current subtopic ID;
  - the current audience ID variable within the security field (which restricts what files are accessible);
  - the start date (set by the administrator) and the end date (current date) range.
- 20

- Group the returned records using the audience ID

(4)      Compare records with the Rating and Usage tables 4

- 25      •      Match audience ID variable with the target field to group content into targeted and general groupings;
- Compare the two groups of library records (targeted and general) with the ratings and usage tables to organize them into a top-to-bottom list with the highest-value content at the top of the list.

30

(5)      Continue loop of Audience table until all records complete

- Set the audience name and ID variables to the next record from the Audience Query
- Repeat steps (3) and (4) for all records returned from the Audience Query

5

(6) Write web page 5

- When all records from the audience query have been processed according to steps (3)-(5), write out the output to a static web page (named as the subtopic ID);
- Add additional security code separating each audience grouping in order to present the proper grouping of information to be displayed when a user accesses the page with a specific audience ID.

10

15

(7) Continue loop of Subtopics table until all records complete

- Continue the above process until all records from the subtopics query have generated a corresponding web page

20

25

The classification-embedded static page is shown in **Figure 1** at 6. Each group corresponds to a record from the Audience query. Within each group, the returned files are organized by targeted content and then general content. And, within each of those groupings, the content from top-to-bottom is stratified as high ratings and high usage (content of the highest value) to content that is of lowest value to the user. Bounding each group is additional code which tells the server which of the groupings of information is to be returned based on the user's specific audience classification when a user accesses the page.

*Using the Classification-Embedded Static Web Pages*

(1) Client and User Settings 21

30

- Each user of the Conjoin Intranet is tracked using information which is stored as a cookie; this information is specific to the

individual user accessing the web site and includes a user's ID, name, audience ID, and preferences or settings.

- At **21**, a user is a member of Group 2 (Audience ID) and has specified that he is interested in subtopics A & D.

5

(2) The Front Page

- When a user accesses the front page of the Conjoin Intranet (or any page in which the CESP pages are used), the front page includes the generated static CESP pages as sub-components of itself. **22**
- At **23**, only the CESP pages that match the user's specific settings are included as part of the Front Page.

10

(3) The Groupings within the CESP page **24**

- Each CESP page has all the possible audience groupings that a user accessing the site could have. The information with the corresponding audience grouping that matches the audience type (or ID) of the user is the one that is displayed.

15

(4) The Front Page : Final Presentation **25**

- The final web page which is displayed to the user is shown schematically at **25**. A typical browser window showing such a final web page is shown in **Figure 4**. Within the page are the CESP pages displaying only the information that is targeted or viewable to the user based on his/her classification and which the user has specified interest in.

20

25

### Exemplary Detailed Database Schema

(a) Content Library Table

- Library Record ID
- Name of Record (assigned by publisher)

30

- 5
- File name (actual name of file or URL from link)
  - Document format (type of file)
  - Author
  - DatePublished
  - DateApproved
  - DateExpired
  - DateArchived
  - Approval\_Status (Yes/No if document is approved to be viewed)
- 10
- Archival\_Status (Yes/No if document is archived)
  - Security (what groups (from Groups Table) have access to see document)
  - Targeting (what groups (from Groups Table) would be most interested in this content item)
- 15
- Taxonomy (what taxonomy records (from Taxonomy Table) is this content item associated with)
- (b) Audience Table
- 20
- Audience Record ID (unique ID)
  - Name
- (c) Subtopics Table
- 25
- The business taxonomy tables contain 2 tables – topics and subtopics. For example, a topic might be competition. The business taxonomy in the subtopics table would include members of competition: Microsoft, Intel, Dell, etc. (see **Figure 5**).
- The taxonomy table contains:
- 30
- Record ID (unique ID)
  - Name

- Category ID (category in which the taxonomy item is a member)

(d) Ratings Table

5 The ratings table contains the individual scores of content which have been rated by users of the application. The ratings table contains the following fields:

- Record ID (unique ID)
- Creation Date (date that the item was created)
- 10 - Audience ID (unique ID from Audience table)
- Content ID (unique ID from library table)
- Rating (score for the item)
- Module (name of module when the item was rated)

15

(e) Usage Table

- Record ID (unique ID)
- Creation date (date that the item was created)
- Audience ID (unique ID from audience table)
- 20 - Content ID (unique ID from library table)
- Module (name of module when the item was rated)

(f) User preference table

- 25 - User ID (unique ID)
- User name (name of the individual user)
- User settings (personal preference data - see **Figure 5**)
- Audience type (audience to which user belongs)

30 Other embodiments of the invention will be apparent to those skilled in the art from a consideration of the specification or practice of the invention disclosed herein.

It is intended that the specification and examples be considered as exemplary only, with the true scope and spirit of the invention being indicated by the following claims.

What is claimed is:

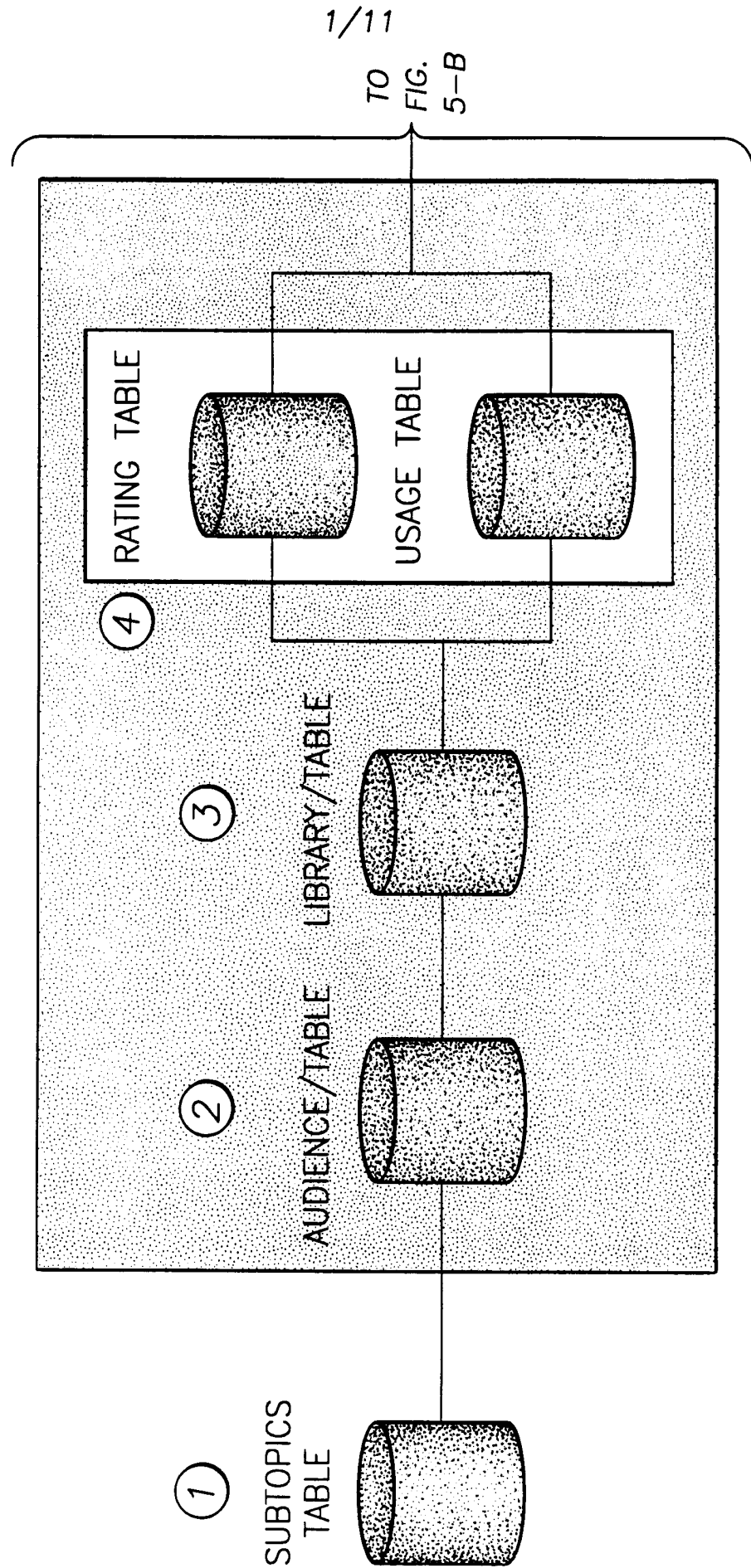
- 1        1.        A computer-implemented method of generating a personalized collection of  
2                    information identifiers, comprising  
3                    maintaining a content table of identifiers for content items, wherein each  
4                    identifier of the content table is classified according to content type and according to  
5                    audience types permitted to access the linked content;  
6                    maintaining a table of users, wherein user has an audience type and at least  
7                    one content interest area;  
8                    generating a plurality of static pages of content identifiers, where each page  
9                    contains identifiers classified as relevant to a single content area, and wherein the  
10                    identifiers of the page are grouped according to the audience access type of their  
11                    associated content items; and  
12                    responding to a request for identifiers for content items from a user by  
13                    selecting the pages related to each of the user's associated content  
14                    areas; and  
15                    returning only those portions of the selected pages corresponding to  
16                    the user's audience type.  
17  
18        2.        The method of claim 1, further comprising sorting a group of identifiers on a  
19                    page associated with a single audience classification according to value of  
20                    their associated content items.  
21  
22        3.        The method of claim 2, wherein the value of the content items is a function of  
23                    the audience type.  
24  
25        4.        The method of claim 1, further comprising adding an additional identifier to  
26                    the content table, where adding includes selecting one or more content types  
27                    and one or more audience access types for the content item.  
28  
29        5.        The method of claim 1, wherein  
30                    each identifier in the content table specifies a set of targeted audience types  
31                    for the associated content item; and

1           the list of identifiers for a single audience access type on a static page is sorted  
2   to place items targeted to that audience type at the beginning of the list.

3

4       6.     The method of claim 1, wherein each identifier of the content database is  
5           selected from the group consisting of uniform resource locators, file location  
6           paths, directories, subdirectories, and catalogue entries for nondigital  
7           information.

FIG. 1-A



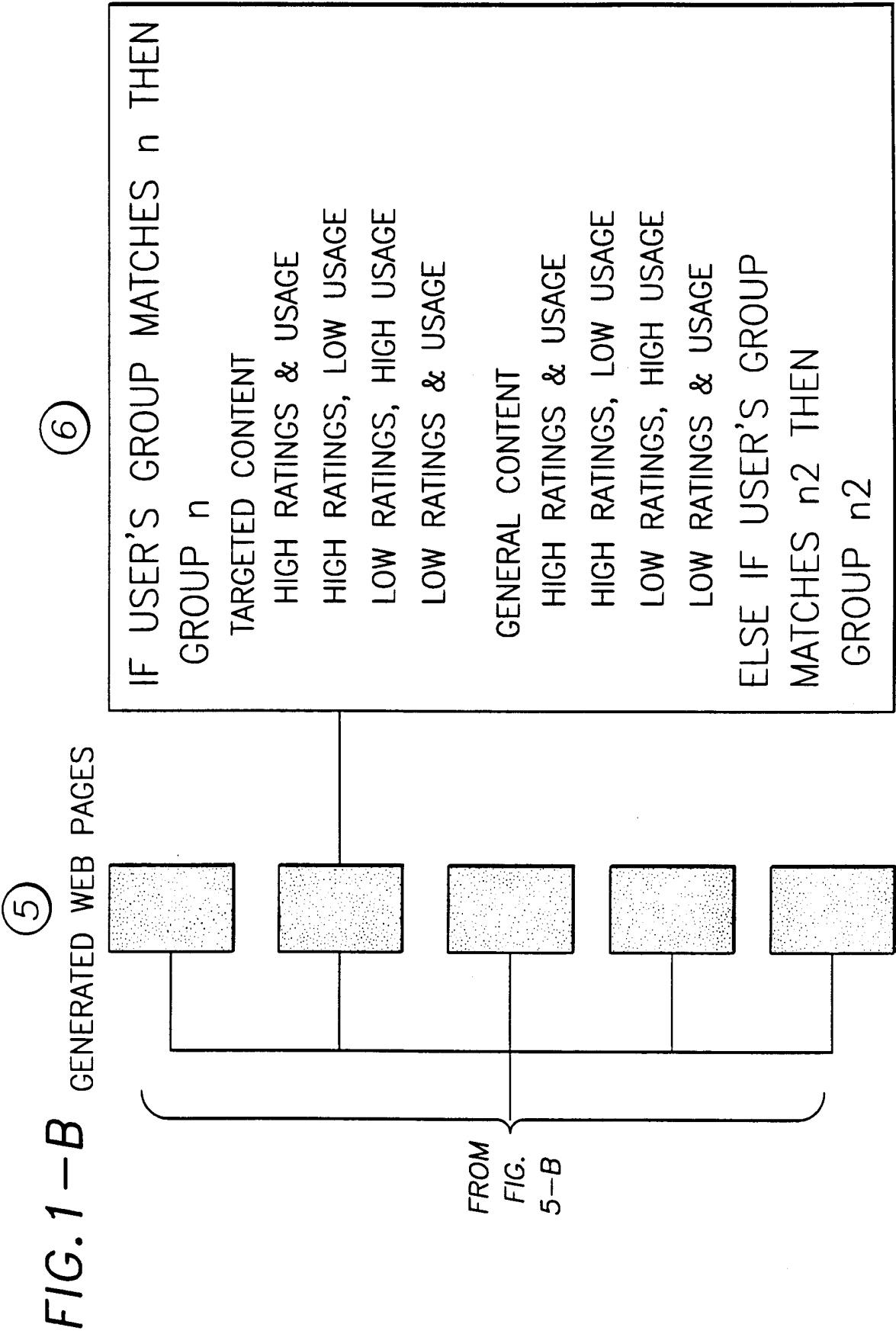


FIG.2

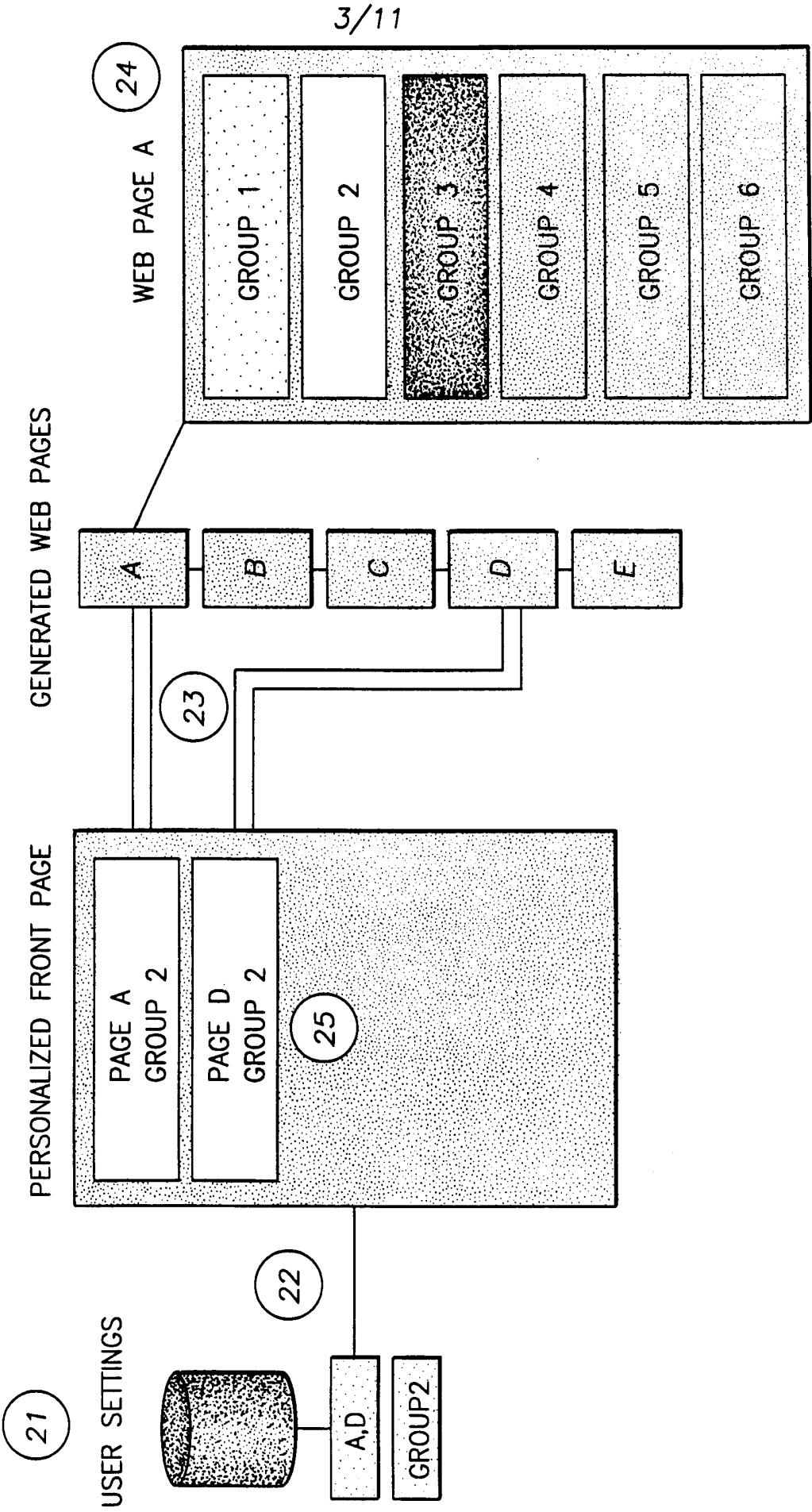
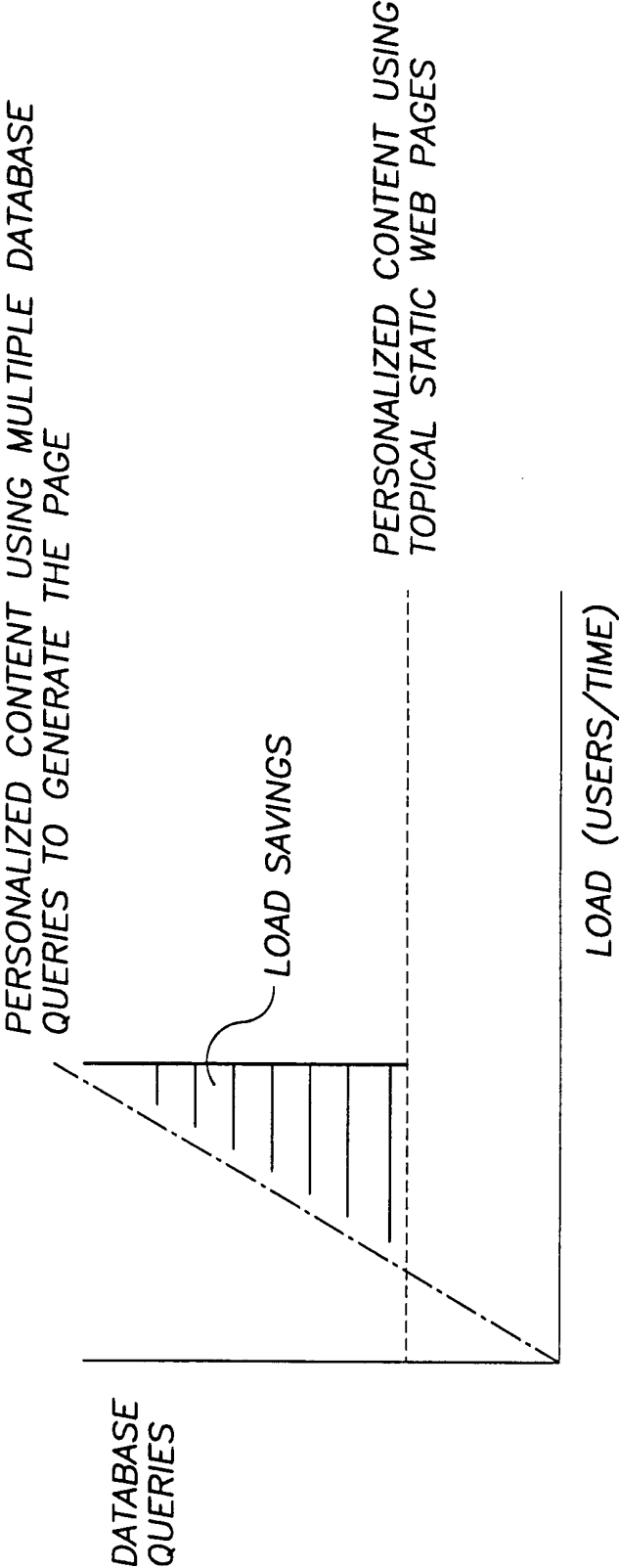


FIG. 3

CONTENT PERSONALIZATION



5/11

FIG. 4-A

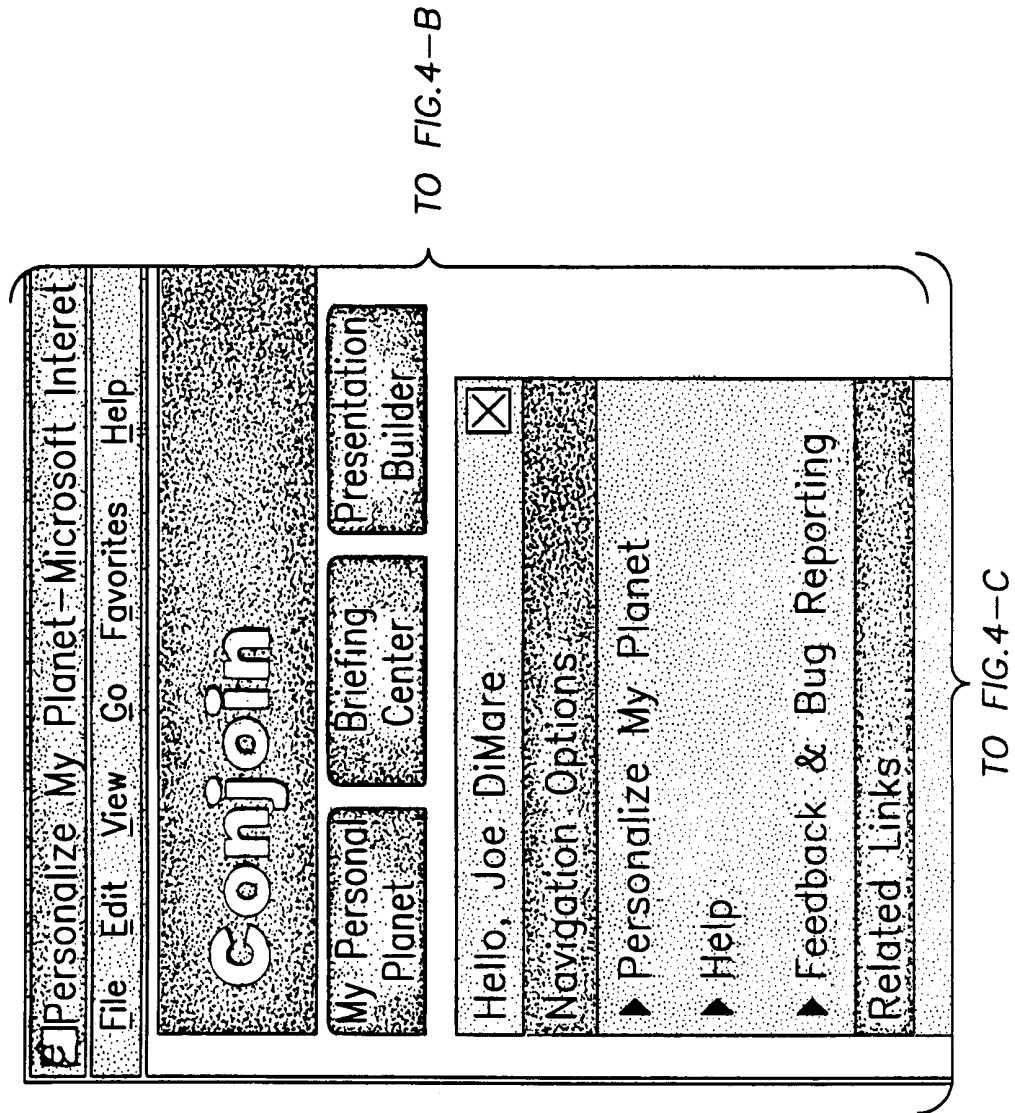
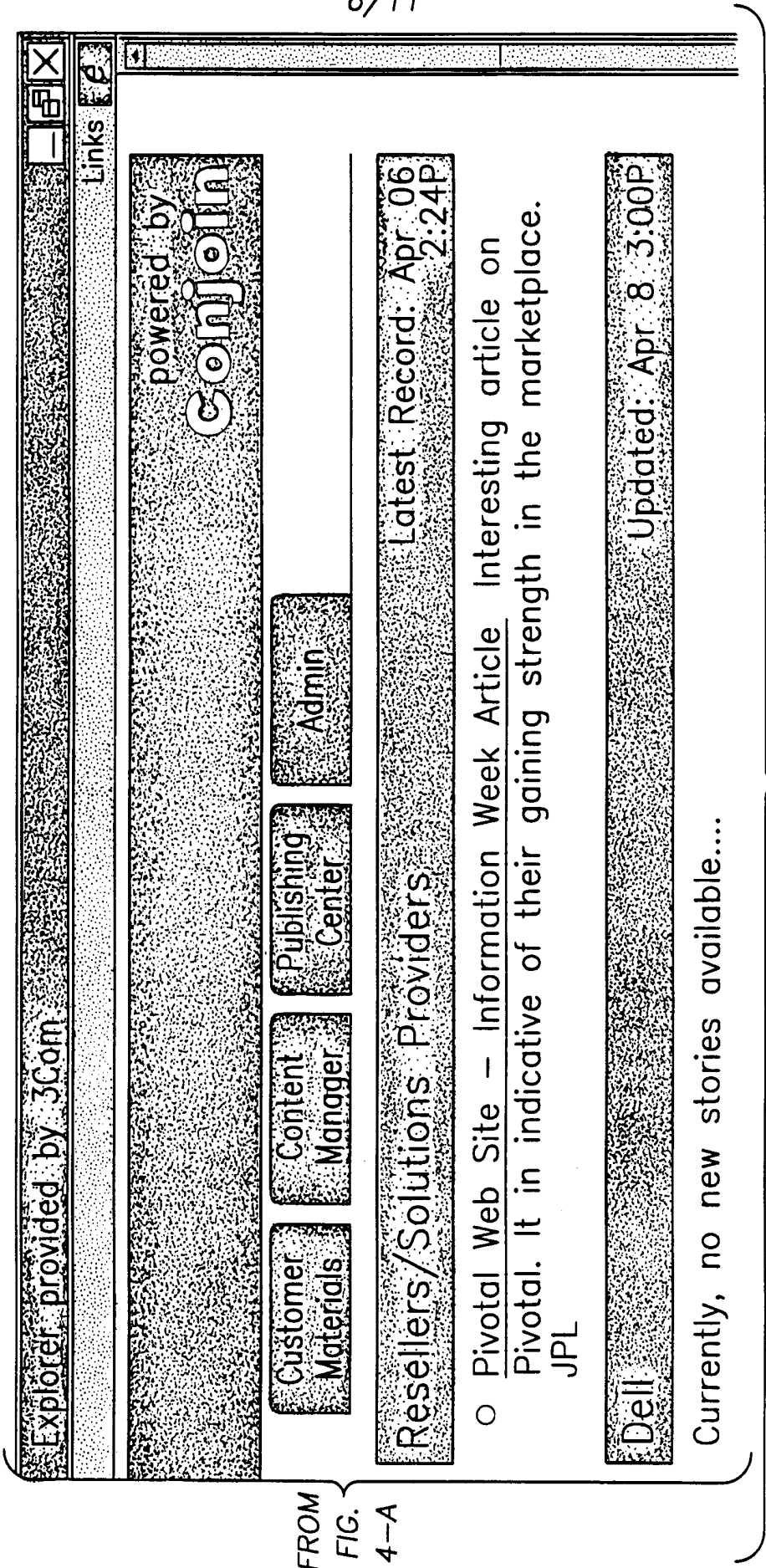


FIG. 4-B

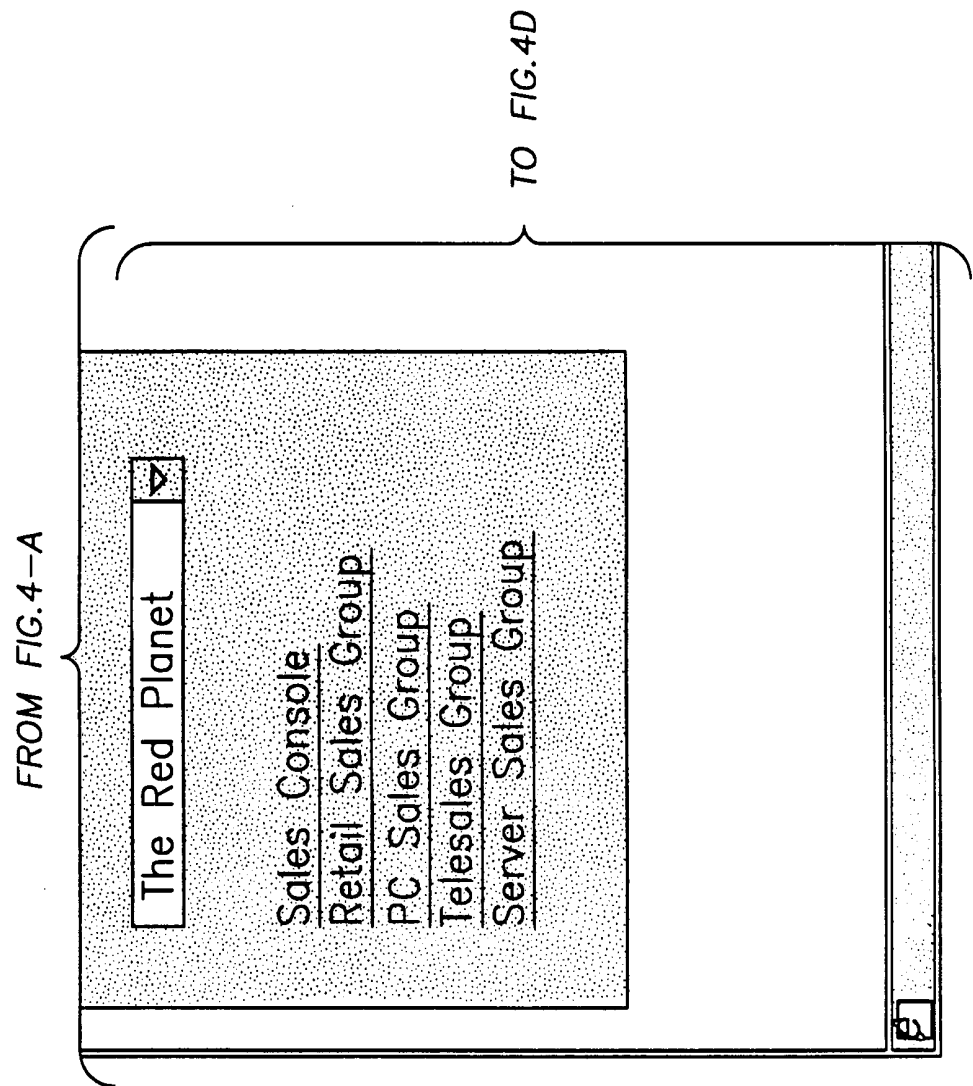


FROM  
FIG.  
4-A

TO FIG. 4-D

7/11

FIG. 4C



8/11

FIG. 4D

FROM FIG. 4-B

Regional Latest Record: Apr 06 3:35P

- NetG's web site NetG has done a good job of creating technology product from their core content base. Check out their site. Patrick
- Presentation to BT Alex Brown this is an Office 95 PPT file. @ 3.3 Megs in size. 36 slides with effects. etc.

OMS Latest Record: Apr 06 2:32P

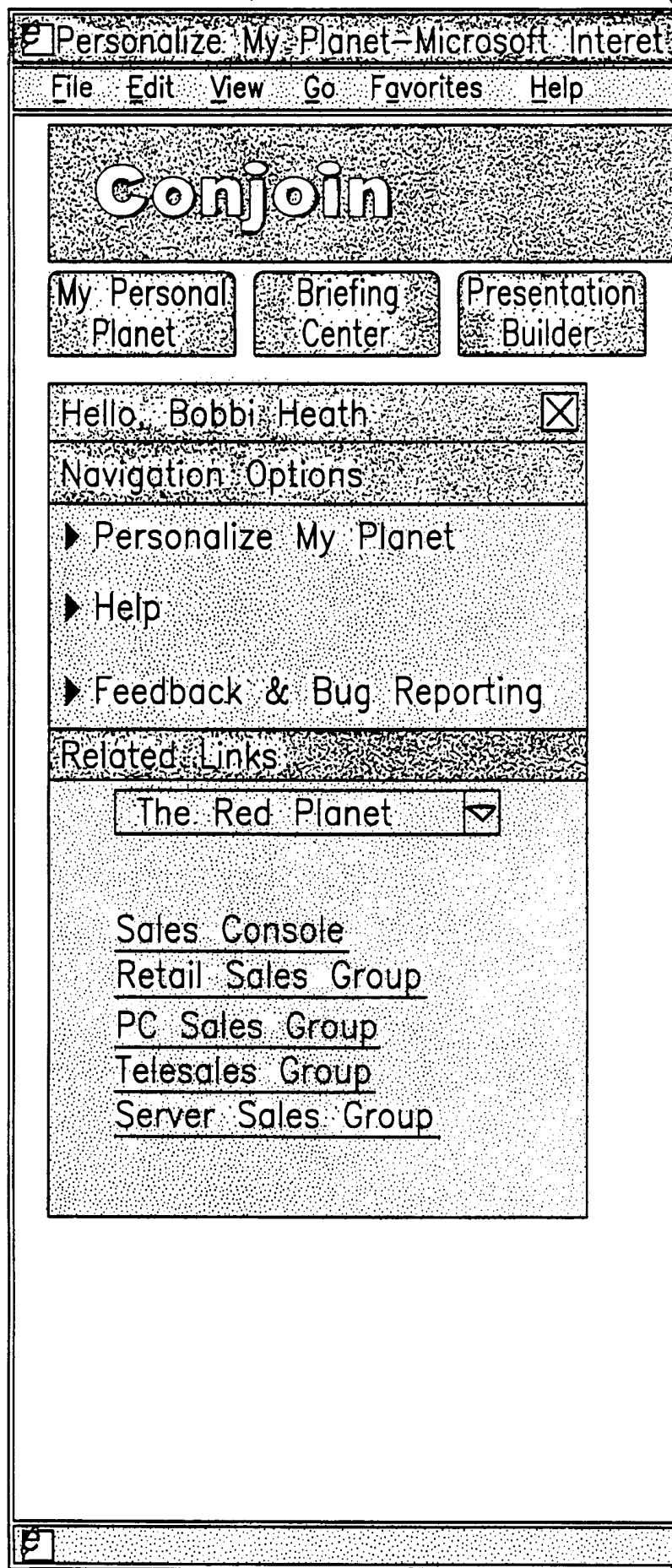
- Ziff Davis PDF Brochure for DC CRM Conference I want to test the indexing of this file by Verity. JPL
- Think Piece on Technology Enabled Revenues Nice piece by Eric Vogt on expanding our revenues. Nice piece by Eric Vogt on expanding our revenues. Nice piece by Eric Vogt on expanding our revenues.

Internetzone

FROM  
FIG.  
4-C

FIG. 5-A

9/11

TO  
FIG.  
5-B

**FIG. 5-B** 10/11

Explorer provided by 3Com

Customer Materials Content Manager Publishing Center Admin

My Personal Planet

☐ Return to My Planet

Personal News  
Select the topics below on which you would like to receive a listing of the latest content:

Channels

- ☐ Distributors
- ☐ ISVs
- ☐ Resellers/Solutions Providers
- ☐ System Integrators/OEMs
- ☐ US1s
- ☐ VARs

Competitors

- ☐ Clarion
- ☒ Dell
- ☐ EMC
- ☐ Gateway
- ☐ HP
- ☐ IBM
- ☐ MTI

FROM  
FIG.  
5-ATO  
FIG.  
5-C

FIG. 5-C

11/11

FROM  
FIG.  
5-B

