NewsHound Database – Final Report

James Reffell
Jean-Anne Fitzpatrick
Moryma Aydelott
IS 257
December 2001
# Table of Contents

Introduction .................................................................................................................. 1
Purpose of the database ............................................................................................... 1
Database description ................................................................................................. 1
  Tables and relations ............................................................................................... 1
  Web sites and pages ............................................................................................... 1
  Geographic data .................................................................................................... 2
  Temporal data ....................................................................................................... 2
Web Application ....................................................................................................... 2
  Data entry approach ............................................................................................. 2
Queries ..................................................................................................................... 4
Forms ...................................................................................................................... 4
Reports .................................................................................................................... 4
Changes made since midterm report ...................................................................... 5
Future work .............................................................................................................. 6

Attachment A. Relationships Diagram
Attachment B. Data Dictionary
Attachment C. PHP Form-Based Queries
Attachment D. MySQL Queries for Data Entry, Update, And Deletion
Attachment E. MySQL Commands for Page Data Import
Attachment F. PHP Data Entry Forms
Attachment G. PHP Reports
Introduction
The goal of the NewsHound project is to create visualizations of themes in web-based news. In particular, we are interested in finding novel ways to visualize changes over time.

In addition, we plan to investigate the differences that emerge between various news sources. For example, we expect there to be striking differences in news coverage from sources based in different geographic locations. For this reason, we are emphasizing geographic as well as temporal metadata.

Purpose of the database
In the short term, the NewsHound database is a medium for exploring the project's content area. In the longer term, we envision integrating the database with an information retrieval system and a web-based user interface providing visualizations. In this context, the database will maintain metadata for the following:

- Web news information sources
- Time ranges for which data is available
- Searchable topics or themes (initially, geographic regions)

Database description
The NewsHound database was initially created in Microsoft Access, but has been converted to MySQL.

Tables and relations
The tables and relationships in the current version of the NewsHound database are shown in the relationships diagram in Attachment A.

A data dictionary for the current version of the NewsHound database is provided in Attachment B.

Web sites and pages
Information about each site from which we are collecting data is stored in the table site. This includes the site's URL, its full name, a description of the site, and a short tag name used for on-going import of page information. The site table also identifies the country of origin and a city of origin where applicable (for instance, for sites associated with local newspapers). A field is provided to indicate the site's political affiliation as described by the World Press Review (http://www.worldpress.org).

For each site, content information is being collected over time using a Perl script which downloads and stores the site HTML. Each snapshot of the site content is identified as a page. The page table contains basic metadata about each collected snapshot: the site ID, the date and time at which the page data was collected, total word count for the page including mark-up, text-only word count, a count of the number of image tags, and the file name in which the HTML is stored.
Geographic data
Several tables provide a hierarchy of geographic data, in terms of regions of the world, countries, and cities. The region table provides a name for each region of the world; the country table provides a name for each country, and includes the region ID as a foreign key; the city table provides a name for each city, and includes the country ID as a foreign key. In addition, the city table includes a flag to indicate if the city is the capital city of its country. The country table also includes fields for population (in thousands) and per capita GDP (in US dollars).

Information about world leaders are stored in the leader table. This includes each leader's name and title, and includes the country ID as as foreign key. At this time, we are not distinguishing between national and local leaders. However, the leader table does include a flag to indicate if the leader is the principle leader of the country. Because leader names are likely to change regularly, this table also includes an explicit field indicating when the information on each leader was updated.

For countries, cities, and leaders, there may be alternative names used. For instance, Beijing may also be referred to as Peking. Such alternate names are stored in tables alt_country, alt_city, and alt_leader.

Temporal data
As noted above, the page table identifies the date and time at which each site content snapshot was obtained. In addition, the holiday table provides an extensive list of days which are celebrated as holidays. Because holidays vary geographically, an intermediate table holiday_country is used to link the holiday ID of each holiday to the country ID of each country where they are celebrated.

Web Application
We are using a PHP-based web interface, which provides access to a number of data entry forms, reports, and queries. Figure 1 shows the main "switchboard" page (http://dream.sims.berkeley.edu/newshound/main.html). Note that in this context we are using the term "report" to refer to a web page that simply displays a predefined set of information, and the term "query" to refer to a web page which uses a user selection to construct the query (both types of pages contain one or more SQL queries).

Data entry approach
Most initial data entry was performed by importing text files into temporary tables and then using queries to populate the database tables; this approach will also be used for on-going import of page information. In addition, web forms created using PHP are available for adding and updating information in several tables. For other tables, updates will be performed directly at the MySQL command line. Also note that at the present time, no forms have been created for deleting items from the database; this will also be performed at the MySQL command line.
Figure 1. NewsHound Database Switchboard Page

NewsHound Database

Available forms, reports, and queries

Forms:
- Enter new country
- Enter or correct data for existing country
- View and enter alternate country names
- Enter new city
- Enter or correct data for existing city
- View and enter alternate city names
- Enter new holiday

Reports:
- Countries by region
- Capital city by country
- Leaders by country
- Sites by region
- Sites by political affiliation

Queries:
- Country profile
- Countries by population and GDP
- Cities by country
- Site profile
- Collected pages
- Where's the party?
- When's the party?
**Queries**

Attachment C provides illustrations and source code for the following PHP query forms:

1. Country profile
2. Countries by population and GDP
3. Cities by country
4. Site profile
5. Collected pages
6. Where's the party?
7. When's the party?

Attachment D shows example queries used for data entry, modification, and deletion for those tables which do not currently have associated forms.

Attachment E describes the process used for page data import, which includes the MySQL `load data` command and an update query.

**Forms**

Attachment F provides illustrations and source code for the following PHP data entry forms:

1. Enter new country
2. Enter or correct data for existing country
3. View and enter alternate country names
4. Enter new city
5. Enter or correct data for existing city
6. View and enter alternate city names
7. Enter new holiday

**Reports**

Attachment G provides illustrations and source code for the following PHP reports:

1. Sites by political affiliation
2. Sites by region
3. Countries by region
4. Capital City by Country
5. Leaders by Country
Changes made since midterm report

Our largest change was the migration of our database from Access to mySQL. This shift provided a chance to examine and confirm our database structure and procedures.

We re-examined all of the data types used, and adjusted several of them.

- In Access, we had deliberately chosen not to use the native date and time types, since we anticipated migrating to another database vendor, and felt that numeric values were more likely to export in a format that would be readily usable. In MySQL, we chose to use the native date and time types. These have proven to work well both in data import (such as the on-going import of page data) and in our PHP-based forms.

- We were also faced with a choice of how to implement boolean fields. In Access, we had used the Yes/No type; however, MySQL does not include a native boolean type. After examining several options, we chose to use an enumerator (enum) type with two values (zero and one). This provides a high degree of data integrity, stores values in a form that is readily interpreted by most application programs, and requires minimal storage space (1 byte).

We added additional fields to tables.

- The country table now includes population (pop) and GDP (gdp).
- The page table now has separate fields to store plain text word count (word_count) as well as the word count including HTML tags (token_count).

We removed unnecessary tables.

- We determined that the scheme we had been using to save link data was generating a large number of entries, but that the data was of no immediate value. On that basis, we decided to remove the link table.

We improved our data collection process.

- As noted above, we now obtain word count values for plain text as well as the full HTML file.
- The scripts used to generate page data for import have been improved, so that the number of errors that occur during the import process is minimized.

We added large amounts of real data.

- We have added more than 3000 cities and 1600 pages to date. This has enabled us to test our database under real-world conditions, and to actively use it for our ongoing project.

We developed mechanisms to allow remote access and facilitate data entry / cleanup.

- We made a substantial effort to develop forms that would be of ongoing use in our project, starting with tables we were actively updating/correcting, and moving on to include tables that we anticipated the need to update in the future.
- Additionally, the development of PHP forms facilitates remote access to the database. We feel this will be important as the project progresses.
Future work

This process of examination, adjustment, and development is ongoing. However, we are aware of some potential weaknesses and therefore currently have plans for work in three major areas.

**Security** - Security has not been a concern up to this point, as our primary concern was developing basic functionality and only the three of us were making changes. However, since the database and procedures involved with it are functional, we are increasingly aware of limitations in our system’s security functionality. Our file index is available to anyone online and our pages don’t require user login before allowing changes. Since we plan to make this system usable by the general public, remedying these and other security problems is a top priority.

**Transactions** - As we began to develop more complex forms for data entry (having already entered in the major portions using import queries), we became aware of the need to maintain transaction integrity. We plan to implement simple locking for cross-table updates and other measures as necessary.

**Connection to an IR system** – We plan on using this database to store metadata, while using an IR system to index and search our collected pages. The work ahead of us includes both design and implementation of the integrated system.
Attachment A: Relationships Diagram
Figure A-1. NewsHound Database Relationships Diagram
Attachment B: Data Dictionary

The following listing provides information on the fields and indices of each of the NewsHound Database tables, as displayed using the standard SQL command `describe` (abbreviated `desc`) and the MySQL command `show index`.

All tables are indexed on their primary key fields (except for page_import, which is used only during the page data import process, and has neither a primary key nor an index).

The describe command lists each field name and type, whether nulls are allowed (YES or blank), whether the field is a key (PRI for primary key or blank), the default value if any, and any "extra" characteristics of the field (for NewsHound, either auto_increment or blank).

The following definitions for the `show index` output are derived from the MySQL web site (http://www.mysql.com):

SHOW INDEX returns the index information in a format that closely resembles the SQLStatistics call in ODBC. The following columns are returned:

<table>
<thead>
<tr>
<th>Column</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table</td>
<td>Name of the table.</td>
</tr>
<tr>
<td>Non_unique</td>
<td>0 if the index can't contain duplicates.</td>
</tr>
<tr>
<td>Key_name</td>
<td>Name of the index.</td>
</tr>
<tr>
<td>Seq_in_index</td>
<td>Column sequence number in index, starting with 1.</td>
</tr>
<tr>
<td>Column_name</td>
<td>Column name.</td>
</tr>
<tr>
<td>Collation</td>
<td>How the column is sorted in the index. In MySQL, this can have values <code>A</code> (Ascending) or NULL (Not sorted).</td>
</tr>
<tr>
<td>Cardinality</td>
<td>Number of unique values in the index. Note that as the Cardinality is counted based on statistics stored as integers, it's not necessarily accurate for small tables.</td>
</tr>
<tr>
<td>Sub_part</td>
<td>Number of indexed characters if the column is only partly indexed. NULL if the entire key is indexed.</td>
</tr>
</tbody>
</table>
mysql> show tables;
+---------------------+
| Tables in newshound |
+---------------------+
| alt_city |
| alt_country |
| alt_leader |
| city |
| country |
| holiday |
| holiday_country |
| leader |
| page |
| page_import |
| politics |
| region |
| site |
+---------------------+
13 rows in set (0.00 sec)
```sql
mysql> desc alt_city;
+----------+-------------+------+-----+---------+-------+
| Field    | Type        | Null | Key  | Default | Extra |
|----------+-------------+------+-----+---------+-------+
| city_ID  | smallint(6) |      | PRI  | 0       |       |
| alt_name | varchar(50) |      | PRI  |         |       |
+----------+-------------+------+-----+---------+-------+
```

2 rows in set (0.00 sec)

```sql
mysql> show index from alt_city;
+----------+------------+----------+--------------+-------------+-----------+-------------+----------+
| Table    | Non_unique | Key_name | Seq_in_index | Column_name | Collation | Cardinality | Sub_part |
|----------+------------+----------+--------------+-------------+-----------+-------------+----------+
| alt_city | 0          | PRIMARY  | 1            | city_ID     | A         | 44          | NULL     |
| alt_city | 0          | PRIMARY  | 2            | alt_name    | A         | 44          | NULL     |
+----------+------------+----------+--------------+-------------+-----------+-------------+----------+
```

2 rows in set (0.00 sec)

```sql
mysql> desc alt_country;
+------------+-------------+------+-----+---------+-------+
| Field      | Type        | Null | Key  | Default | Extra |
|------------+-------------+------+-----+---------+-------+
| country_ID | smallint(6) |      | PRI  | 0       |       |
| alt_name   | varchar(50) |      | PRI  |         |       |
+------------+-------------+------+-----+---------+-------+
```

2 rows in set (0.00 sec)

```sql
mysql> show index from alt_country;
+-------------+------------+----------+--------------+-------------+-----------+-------------+----------+
| Table       | Non_unique | Key_name | Seq_in_index | Column_name | Collation | Cardinality | Sub_part |
|-------------+------------+----------+--------------+-------------+-----------+-------------+----------+
| alt_country | 0          | PRIMARY  | 1            | country_ID  | A         | 60          | NULL     |
| alt_country | 0          | PRIMARY  | 2            | alt_name    | A         | 60          | NULL     |
+-------------+------------+----------+--------------+-------------+-----------+-------------+----------+
```

2 rows in set (0.00 sec)
mysql> desc alt_leader;

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Null</th>
<th>Key</th>
<th>Default</th>
<th>Extra</th>
</tr>
</thead>
<tbody>
<tr>
<td>leader_ID</td>
<td>smallint(6)</td>
<td>0</td>
<td>PRI</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>alt_name</td>
<td>varchar(50)</td>
<td>PRI</td>
<td>PRI</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2 rows in set (0.00 sec)

mysql> show index from alt_leader;

<table>
<thead>
<tr>
<th>Table</th>
<th>Non_unique</th>
<th>Key_name</th>
<th>Seq_in_index</th>
<th>Column_name</th>
<th>Collation</th>
<th>Cardinality</th>
<th>Sub_part</th>
</tr>
</thead>
<tbody>
<tr>
<td>alt_leader</td>
<td>0</td>
<td>PRIMARY</td>
<td>1</td>
<td>leader_ID</td>
<td>A</td>
<td>6</td>
<td>NULL</td>
</tr>
<tr>
<td>alt_leader</td>
<td>0</td>
<td>PRIMARY</td>
<td>2</td>
<td>alt_name</td>
<td>A</td>
<td>6</td>
<td>NULL</td>
</tr>
</tbody>
</table>

2 rows in set (0.00 sec)

mysql> desc city;

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Null</th>
<th>Key</th>
<th>Default</th>
<th>Extra</th>
</tr>
</thead>
<tbody>
<tr>
<td>city_ID</td>
<td>smallint(6)</td>
<td>0</td>
<td>PRI</td>
<td>0</td>
<td>auto_increment</td>
</tr>
<tr>
<td>country_ID</td>
<td>smallint(6)</td>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>city_name</td>
<td>varchar(50)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cap_city</td>
<td>enum('0','1')</td>
<td>YES</td>
<td></td>
<td>NULL</td>
<td></td>
</tr>
</tbody>
</table>

4 rows in set (0.00 sec)

mysql> show index from city;

<table>
<thead>
<tr>
<th>Table</th>
<th>Non_unique</th>
<th>Key_name</th>
<th>Seq_in_index</th>
<th>Column_name</th>
<th>Collation</th>
<th>Cardinality</th>
<th>Sub_part</th>
</tr>
</thead>
<tbody>
<tr>
<td>city</td>
<td>0</td>
<td>PRIMARY</td>
<td>1</td>
<td>city_ID</td>
<td>A</td>
<td>3003</td>
<td>NULL</td>
</tr>
</tbody>
</table>

1 row in set (0.00 sec)
mysql> desc country;
+--------------+--------------+------+-----+---------+----------------+
| Field | Type | Null | Key | Default | Extra |
+--------------+--------------+------+-----+---------+----------------+
| country_ID | smallint(6) | | PRI | 0 | auto_increment |
| country_name | varchar(50) | | | | |
| region_ID | smallint(6) | | | 0 | |
| gdp | mediumint(9) | YES | | NULL | |
| pop | mediumint(9) | YES | | NULL | |
+--------------+--------------+------+-----+---------+----------------+
5 rows in set (0.00 sec)

mysql> show index from country;
+---------+------------+----------+--------------+-------------+-----------+-------------+----------+
| Table | Non_unique | Key_name | Seq_in_index | Column_name | Collation | Cardinality | Sub_part |
| country | 0 | PRIMARY | 1 | country_ID | A | 192 | NULL |
+---------+------------+----------+--------------+-------------+-----------+-------------+----------+
1 row in set (0.00 sec)

mysql> desc holiday;
+--------------+--------------+------+-----+------------+----------------+
| Field | Type | Null | Key | Default | Extra |
+--------------+--------------+------+-----+------------+----------------+
| holiday_ID | smallint(6) | | PRI | 0 | auto_increment |
| date | date | | 0000-00-00 | |
| holiday_name | varchar(255) | | | | |
+--------------+--------------+------+-----+------------+----------------+
3 rows in set (0.00 sec)

mysql> show index from holiday;
+---------+------------+----------+--------------+-------------+-----------+-------------+----------+
| Table | Non_unique | Key_name | Seq_in_index | Column_name | Collation | Cardinality | Sub_part |
| holiday | 0 | PRIMARY | 1 | holiday_ID | A | 622 | NULL |
+---------+------------+----------+--------------+-------------+-----------+-------------+----------+
1 row in set (0.00 sec)
mysql> desc holiday_country;
+------------+-------------+------+-----+---------+-------+
| Field      | Type        | Null | Key | Default | Extra |
|------------+-------------+------+-----+---------+-------|
| country_ID | smallint(6) |      | PRI | 0       |       |
| holiday_ID | smallint(6) |      | PRI | 0       |       |
+------------+-------------+------+-----+---------+-------+
2 rows in set (0.00 sec)

mysql> show index from holiday_country;
+-----------------+------------+----------+--------------+-------------+-----------+-------------+----------+
| Table           | Non_unique | Key_name | Seq_in_index | Column_name | Collation | Cardinality | Sub_part |
|-----------------+------------+----------+--------------+-------------+-----------+-------------+----------|
| holiday_country | 0          | PRIMARY  | 1            | country_ID  | A         | 1218        | NULL     |
| holiday_country | 0          | PRIMARY  | 2            | holiday_ID  | A         | 1218        | NULL     |
+-----------------+------------+----------+--------------+-------------+-----------+-------------+----------+
2 rows in set (0.00 sec)

mysql> desc leader;
+--------------+---------------+------+-----+------------+----------------+|
| Field        | Type          | Null | Key | Default    | Extra          |
|--------------+---------------+------|-----+------------+----------------+|
| leader_ID    | smallint(6)   |      | PRI | 0          | auto_increment |
| country_ID   | smallint(6)   |      |     | 0          |                |
| date_updated | date          |      |     | 0000-00-00 |                |
| leader_name  | varchar(255)  |      |     |            |                |
| leader_title | varchar(255)  |      |     |            |                |
| big_boss     | enum('0','1') | YES  |     | NULL       |                |
+--------------+---------------+------|-----+------------+----------------+
6 rows in set (0.00 sec)

mysql> show index from leader;
+--------+------------+----------+--------------+-------------+-----------+-------------+----------+
| Table  | Non_unique | Key_name | Seq_in_index | Column_name | Collation | Cardinality | Sub_part |
|--------+------------+----------+--------------+-------------+-----------+-------------+----------|
| leader | 0          | PRIMARY  | 1            | leader_ID   | A         | 181         | NULL     |
+--------+------------+----------+--------------+-------------+-----------+-------------+----------+
1 row in set (0.00 sec)
mysql> desc page;
+----------------+-------------+------+-----+------------+----------------+
| Field          | Type        | Null | Key | Default    | Extra          |
+----------------+-------------+------+-----+------------+----------------+
| page_id        | smallint(6) |      | PRI | 0          | auto_increment |
| site_ID        | smallint(6) |      |     | 0          |                |
| date           | date        |      |     | 0000-00-00 |                |
| time           | time        |      |     | 00:00:00   |                |
| word_count     | smallint(6) |      |     | 0          |                |
| full_text_file | varchar(50) |      |     |            |                |
| image_count    | smallint(6) |      |     | 0          |                |
| token_count    | smallint(6) |      |     | 0          |                |
+----------------+-------------+------+-----+------------+----------------+
8 rows in set (0.00 sec)

mysql> show index from page;
+-------+------------+----------+--------------+-------------+-----------+-------------+----------+
| Table | Non_unique | Key_name | Seq_in_index | Column_name | Collation | Cardinality | Sub_part |
+-------+------------+----------+--------------+-------------+-----------+-------------+----------+
| page  | 0          | PRIMARY  | 1            | page_id     | A         | 1721        | NULL     |
+-------+------------+----------+--------------+-------------+-----------+-------------+----------+
1 row in set (0.00 sec)

mysql> desc page_import;
+----------------+-------------+------+-----+---------+-------+
| Field          | Type        | Null | Key | Default | Extra |
+----------------+-------------+------+-----+---------+-------+
| date           | date        | YES  |     | NULL    | NULL  |
| time           | time        | YES  |     | NULL    | NULL  |
| word_count     | smallint(6) | YES  |     | NULL    | NULL  |
| full_text_file | varchar(50) | YES  |     | NULL    | NULL  |
| image_count    | smallint(6) | YES  |     | NULL    | NULL  |
| site_tag       | varchar(20) | YES  |     | NULL    | NULL  |
| token_count    | smallint(6) | YES  |     | NULL    | NULL  |
+----------------+-------------+------+-----+---------+-------+
7 rows in set (0.00 sec)
mysql> desc politics;
+-------------+-------------+------+-----+---------+----------------+
| Field       | Type        | Null | Key | Default | Extra          |
+-------------+-------------+------+-----+---------+----------------+
| politics_ID | smallint(6) |      | PRI | 0       | auto_increment |
| politics    | varchar(50) |      |     |         |                |
+-------------+-------------+------+-----+---------+----------------+
2 rows in set (0.00 sec)

mysql> show index from politics;
+----------+------------+----------+--------------+-------------+-----------+-------------+----------+
| Table    | Non_unique | Key_name | Seq_in_index | Column_name | Collation | Cardinality | Sub_part |
+----------+------------+----------+--------------+-------------+-----------+-------------+----------+
| politics | 0          | PRIMARY  | 1            | politics_ID | A         | 7           | NULL     |
+----------+------------+----------+--------------+-------------+-----------+-------------+----------+
1 row in set (0.00 sec)

mysql> desc region;
+-------------+-------------+------+-----+---------+----------------+
| Field       | Type        | Null | Key | Default | Extra          |
+-------------+-------------+------+-----+---------+----------------+
| region_ID   | smallint(6) |      | PRI | 0       | auto_increment |
| region_name | varchar(50) |      |     |         |                |
+-------------+-------------+------+-----+---------+----------------+
2 rows in set (0.00 sec)

mysql> show index from region;
+--------+------------+----------+--------------+-------------+-----------+-------------+----------+
| Table  | Non_unique | Key_name | Seq_in_index | Column_name | Collation | Cardinality | Sub_part |
+--------+------------+----------+--------------+-------------+-----------+-------------+----------+
| region | 0          | PRIMARY  | 1            | region_ID   | A         | 7           | NULL     |
+--------+------------+----------+--------------+-------------+-----------+-------------+----------+
1 row in set (0.00 sec)
mysql> desc site;

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Null</th>
<th>Key</th>
<th>Default</th>
<th>Extra</th>
</tr>
</thead>
<tbody>
<tr>
<td>site_ID</td>
<td>smallint(6)</td>
<td></td>
<td>PRI</td>
<td>0</td>
<td>auto_increment</td>
</tr>
<tr>
<td>site_tag</td>
<td>varchar(20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>site_name</td>
<td>varchar(50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>site_description</td>
<td>varchar(255)</td>
<td>YES</td>
<td>NULL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>site_URL</td>
<td>varchar(255)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>country_ID</td>
<td>smallint(6)</td>
<td></td>
<td>YES</td>
<td>0</td>
<td>NULL</td>
</tr>
<tr>
<td>city_ID</td>
<td>smallint(6)</td>
<td></td>
<td>YES</td>
<td>NULL</td>
<td></td>
</tr>
<tr>
<td>politics_ID</td>
<td>smallint(6)</td>
<td></td>
<td>YES</td>
<td>NULL</td>
<td></td>
</tr>
</tbody>
</table>

8 rows in set (0.01 sec)

mysql> show index from site;

<table>
<thead>
<tr>
<th>Table</th>
<th>Non_unique</th>
<th>Key_name</th>
<th>Seq_in_index</th>
<th>Column_name</th>
<th>Collation</th>
<th>Cardinality</th>
<th>Sub_part</th>
</tr>
</thead>
<tbody>
<tr>
<td>site</td>
<td>0</td>
<td>PRIMARY</td>
<td>1</td>
<td>site_ID</td>
<td>A</td>
<td>58</td>
<td>NULL</td>
</tr>
</tbody>
</table>

1 row in set (0.00 sec)
Attachment C: PHP Form-based Queries

The following pages provide illustrations and source code for the following PHP-based queries:

1. Country profile
   (http://dream.sims.berkeley.edu/newshound/country_profile.php)

2. Countries by population and GDP
   (http://dream.sims.berkeley.edu/newshound/country_pop_gdp.php)

3. Cities by country
   (http://dream.sims.berkeley.edu/newshound/city_by_country.php)

4. Site profile
   (http://dream.sims.berkeley.edu/newshound/site_profile.php)

5. Collected pages
   (http://dream.sims.berkeley.edu/newshound/pages_by_date.php)

6. Where's the party?
   (http://dream.sims.berkeley.edu/newshound/holiday.php)

7. When's the party?
   (http://dream.sims.berkeley.edu/newshound/country_holiday.php)

Queries are highlighted using a shaded box.
1. Country profile
   (http://dream.sims.berkeley.edu/newshound/country_profile.php)
Source file country_profile.php

```php
<html><head>
<title>NewsHound Database Query - Country Profile</title>
<link rel=stylesheet type="text/css" href="styles.css">
</head>

<body bgcolor="#ffffff">
<h1><a href="main.html"><img src="clearhound.gif" border=0></a>&nbsp;NewsHound</h1><hr>
<h2>Country Profile</h2>

```

```php
mysql_connect("localhost", "newshound", "newshound");
if (isset($country_id) && $country_id!="") {

# get currently selected country name
$q_where = "SELECT country_name
        from country
        where country.country_id = $country_id";
$r_where = mysql_db_query("newshound", $q_where);

if (mysql_num_rows($r_where)>1) {
    echo "DEBUG error in query selecting name labels";
}
$r = mysql_fetch_array($r_where);
$thiscountry = $r["country_name"];  
$thiscountry_id = $country_id;

}

echo "<form action=$PHP_SELF method=POST>
<table>
<tr>
<td align=right>
Select Country: 
</td><td>
<select name=country_id>
$q_c = "SELECT country_id, country_name
        FROM country
        ORDER BY country_name";
$r_c = mysql_db_query("newshound", $q_c);

if ($r_c) {
    while ($r = mysql_fetch_array($r_c)) {
        $country_id = $r["country_id"];  
        $country_name = $r["country_name"];  
        if ($country_name==$thiscountry) {
            echo "<option selected value=$country_id>$country_name";
        }
    }
```
else {
    echo "<option value=$country_id>$country_name"
}
}
}

echo "</select><br>";

$country_id = $thiscountry_id;

if (isset($country_id) && $country_id!="") {
    # get currently selected country name
    $q_where = "SELECT country_name
    from country
    where country.country_id = $country_id";

    $r_where = mysql_db_query("newshound", $q_where);
    if (mysql_num_rows($r_where)>1) {
        echo "DEBUG error in query selecting name labels"
    }
    $r = mysql_fetch_array($r_where);
    $thiscountry = $r["country_name"];

    if (isset($check)) {
        $q_check = "select country_name, region_name,
        gdp, pop, leader_name
        from country, region left outer join leader
        on leader.country_ID = country.country_ID
        and big_boss = '1'
        where country.region_ID = region.region_ID
        and country.country_ID = $country_id";

        $r_check = mysql_db_query("newshound", $q_check);
        if ( $r_check && mysql_num_rows($r_check)>0 ) {
            echo "<p>Profile for $thiscountry: <p>"
            echo "<p<table width = '90%'><n<tr><n
            <td>Country</td><td>Region</td>
            <td>GDP (per capita)</td>
            <td>Population</td><td>Primary Leader</td></tr>
            <td>$country_name</td>
            <td>$region_name</td>
            <td>$GDP</td>
            <td>$population</td>
            <td>$leader_name</td>
            </table></n";
        }
    }
}
else {
    echo "<option value=$country_id>$country_name"
}
}
while ($r = mysql_fetch_array($r_check)) {
    echo "<tr>";
    $country_name = $r["country_name"];  
    $region_name = $r["region_name"]; 
    $gdp = number_format($r["gdp"]); 
    $pop = number_format($r["pop"]); 
    $leader = $r["leader_name"]; 
    echo "<td>$country_name</td><td>$region_name</td>"; 
    if ($gdp!="0") {echo "<td>\$$gdp</td>";}
    else {echo "<td align=center> - </td>";}
    if ($pop!="0") {echo "<td>$pop,000</td>";}
    else {echo "<td align=center> - </td>";}
    if ($leader!="" ) {echo "<td>$leader</td>";}
    else {echo "<td align=center> - </td>";}
    echo "</tr>";
}

else {
    echo "<p>Alert! No profile found for $thiscountry";
}
} # end if check current entries

} # end isset country id

if ($r_c) {
    mysql_free_result($r_c);
}

?>

<p><img src="smclearhound.gif">&nbsp;&nbsp;<a href="main.html">Back to main NewsHound page</a></p>

</body>

</html>
2. Countries by population and GDP
   (http://dream.sims.berkeley.edu/newshound/country_pop_gdp.php)

<table>
<thead>
<tr>
<th>Country</th>
<th>Population</th>
<th>GDP (US$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td></td>
<td>$1,726</td>
</tr>
<tr>
<td>TestCountry</td>
<td>900,000</td>
<td>$72</td>
</tr>
<tr>
<td>Israel</td>
<td>3,047,000</td>
<td>$17,564</td>
</tr>
<tr>
<td>Bolivia</td>
<td>4,239,000</td>
<td>$1,032</td>
</tr>
<tr>
<td>Jordan</td>
<td>5,051,000</td>
<td>$1,576</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>7,960,000</td>
<td>$23,579</td>
</tr>
<tr>
<td>Chile</td>
<td>15,401,000</td>
<td>$4,505</td>
</tr>
<tr>
<td>Australia</td>
<td>19,339,000</td>
<td>$21,319</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>21,028,000</td>
<td>$7,095</td>
</tr>
<tr>
<td>Afghanistan</td>
<td>22,474,000</td>
<td>$178</td>
</tr>
<tr>
<td>Canada</td>
<td>31,015,000</td>
<td>$20,822</td>
</tr>
<tr>
<td>Kenya</td>
<td>31,293,000</td>
<td>$355</td>
</tr>
</tbody>
</table>
Source file country_pop_gdp.php

<html>
<head>
<title>NewsHound Database Query - Population and GDP by Country</title>
<link rel=stylesheet type="text/css" href="styles.css">
</head>

<body>
<a name=top></a>
<h1><a href="main.html"><img src="clearhound.gif" border=0></a>&nbsp;NewsHound</h1><hr>
<h2>Population and GDP by Country</h2><p><a href="#b">Jump to bottom of list</a></p>
<?
# the arguments to mysql_connect are hostname, username, and password
mysql_connect("localhost", "newshound", "newshound");

# main data used in page
$q_country = "SELECT DISTINCT country_name, pop, gdp 
FROM country"

if (isset($bypop)) {
$q_country .= " ORDER BY pop";
} else if (isset($bygdp)) {
$q_country .= " ORDER BY gdp";
} else {
$q_country .= " ORDER BY country_name"
}

# echo "<br>DEBUG query is $q_country"

# the arguments to mysql_db_query are dbname, and query string
$r_country = mysql_db_query("newshound", $q_country);
if ($r_country) {
  echo "<form action=$PHP_SELF method=POST>
  echo "<table cellspacing=5 cellpadding=5>"
  echo "<tr>"
  echo "<td><input type=submit name='bycountry' value='Sort by country name'></td>
  echo "<td colspan=2><input type=submit name='bypop' value='Sort by population'></td>
  echo "<td colspan=2><input type=submit name='bygdp' value='Sort by GDP (US$)'></td>
  echo "</tr>";
while ($r = mysql_fetch_array($r_country)) {
    $country_name = $r["country_name"];  
    $pop = $r["pop"]; 
    $pop = number_format($r["pop"]); 
    $gdp = number_format($r["gdp"]); 
    if ($gdp!=0 || $pop!=0) {
        if ($pop=="0") {
            $pop = "-";
        } else {
            $pop .= ",000";
        }
        if ($gdp=="0") {
            $gdp = "-";
        } else {
            $gdp = "$".$gdp;
        }
        echo "<tr>";
        echo "<td>$country_name</td>";
        echo "<td align=right>$pop</td><td>&nbsp;</td>";
        echo "<td align=right>$gdp</td><td>&nbsp;</td>";
        echo "</tr>";
    }
}

echo "</table>";
echo "<a name=b></a>";
mysql_free_result($r_country);
} else {
    echo "No data.";
}

<a href=#top>Return to top</a>
<p></p><hr><p><img src="smclearhound.gif">&nbsp;<a href="main.html">Back to main NewsHound page</a>
</body>
</html>
3. Cities by country
(http://dream.sims.berkeley.edu/newshound/city_by_country.php)
### Source file city_by_country.php

```php
<html>
<head>
    <title>NewsHound Database Query - Cities by Country</title>
    <link rel=stylesheet type="text/css" href="styles.css">
</head>

<body bgcolor=#ffffff>
    <a name=top></a>
    <h1><a href="main.html"><img src="clearhound.gif" border=0></a>&nbsp;NewsHound</h1>
    <hr>
    <h2>Cities by Country</h2>
    <p>
        mysql_connect("localhost", "newshound", "newshound");
        <br>
        if (isset($sel_country)) {
            # get currently selected country name
            $q_where = "SELECT country_name
                          from country
                          where country.country_id = $sel_country";
            $r_where = mysql_db_query("newshound", $q_where);
            if (mysql_num_rows($r_where)>1) {
                echo "DEBUG error in query selecting name labels";
            }
            $r = mysql_fetch_array($r_where);
            $thiscountry = $r["country_name"];
            $thiscountry_id = $sel_country;
        }
        echo "<form action=$PHP_SELF method=POST>
        <table>
            <tr>
                <td align=right>
                    Select Country: 
                </td><td>
                <select name=sel_country>
                    $q_c = "SELECT country_id, country_name
                          FROM country
                          ORDER BY country_name";
                    $r_c = mysql_db_query("newshound", $q_c);
                    if ($r_c) {
                        while ($r = mysql_fetch_array($r_c)) {
                            $country_id = $r["country_id"];
                            $country_name = $r["country_name"];
                            if ($country_name==$thiscountry) {
                                echo "<option selected value=$country_id>$country_name",
                            }
                        }
                    }
    </form>
</body>
</html>
```
else {
    echo "<option value=$country_id>$country_name";
}
}
}

else {
    echo "</select><br>";
    echo "</td><td>
    echo "<input type=submit name='check' value='List Cities'>";
    echo "</td><tr><td>
    echo "</table>
    echo "</form>";
    if (isset($sel_country)) {
        if (isset($check)) {

            $q_check = "select city_name, cap_city from city
            where city.country_id = $sel_country
            order by city_name";

            $r_check = mysql_db_query("newshound", $q_check);

            if ( $r_check && mysql_num_rows($r_check)>0 ) {
                echo "<p class=bold>Cities in $thiscountry: </p>";
                echo "<ul>";
                while ($r = mysql_fetch_array($r_check)) {
                    $city_name = $r["city_name"];
                    $cap_city = $r["cap_city"];
                    echo "<li>$city_name</li>";
                }
                echo "</ul>";
                echo "<p>
                echo "<a href=#top>Return to top</a>";
            } else {
                echo "<p>No cities were found for $thiscountry";
            }
        } # end if check
    } # end if isset c id

    if ($r_c) {
        mysql_free_result($r_c);
    }
}

?>

<hr><p><img src="smclearhound.gif">&nbsp;<a href="main.html">Back to main NewsHound page</a>

</body>
</html>
4. Site profile
(http://dream.sims.berkeley.edu/newshound/site_profile.php)

<table>
<thead>
<tr>
<th>Site Name</th>
<th>Token Count</th>
<th>Word Count</th>
<th>Image Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCNews.com</td>
<td>3260</td>
<td>438</td>
<td>84</td>
</tr>
<tr>
<td>allAFRICA.com</td>
<td>3806</td>
<td>776</td>
<td>152</td>
</tr>
<tr>
<td>BBC News</td>
<td>3441</td>
<td>803</td>
<td>170</td>
</tr>
<tr>
<td>CBSNews.com</td>
<td>5458</td>
<td>803</td>
<td>57</td>
</tr>
<tr>
<td>CNN.com</td>
<td>2384</td>
<td>418</td>
<td>53</td>
</tr>
<tr>
<td>The Detroit News.com</td>
<td>3977</td>
<td>782</td>
<td>176</td>
</tr>
<tr>
<td>Environmental News Network</td>
<td>2670</td>
<td>584</td>
<td>112</td>
</tr>
<tr>
<td>Fox News Channel</td>
<td>3026</td>
<td>668</td>
<td>252</td>
</tr>
<tr>
<td>The Los Angeles Times</td>
<td>5434</td>
<td>1241</td>
<td>187</td>
</tr>
<tr>
<td>MSNBC</td>
<td>1168</td>
<td>268</td>
<td>24</td>
</tr>
<tr>
<td>Pittsburgh Post-Gazette</td>
<td>2310</td>
<td>804</td>
<td>102</td>
</tr>
<tr>
<td>The San Francisco Examiner</td>
<td>1612</td>
<td>515</td>
<td>56</td>
</tr>
<tr>
<td>USA Today</td>
<td>3045</td>
<td>861</td>
<td>113</td>
</tr>
<tr>
<td>USNews.com</td>
<td>2881</td>
<td>700</td>
<td>70</td>
</tr>
<tr>
<td>The Washington Times</td>
<td>2070</td>
<td>615</td>
<td>60</td>
</tr>
</tbody>
</table>
<?
mysql_connect("localhost", "newshound", "newshound");
if (isset($sel_site)) {
    # get currently selected country name
    $q_where = "SELECT site_name
        from site
        where site_id = $sel_site";
    $r_where = mysql_db_query("newshound", $q_where);
    if (mysql_num_rows($r_where)>1) {
        echo "DEBUG error in query selecting name labels";
    }
    $r = mysql_fetch_array($r_where);
    $thissite = $r["site_name"];
    $thissite_id = $sel_site;
}

echo "<form action=$PHP_SELF method=POST>
    <table>
    <tr>
        <td align=right>
            Select Country: 
        </td><td>
    <select name=sel_site>
$q_sites = "SELECT site_id, site_name
    FROM site
    ORDER BY site_name";
$r_sites = mysql_db_query("newshound", $q_sites);
if ($r_sites) {
    while ($r = mysql_fetch_array($r_sites)) {
        $site_id = $r["site_id"];
        $site_name = $r["site_name"];
        if ($site_name==$thissite) {
            echo "<option selected value=$site_id>$site_name"
        }
    }
}
```php
else {
    echo "<option value=".$site_id."$site_name";
}
}

echo "</select><br>";

echo "</td></tr><tr><td>
   echo "&nbsp;";
   echo "</td><td><input type=submit name='check' value='Check site profile'>";
   echo "</td></tr><tr><td>
   echo "</table>";
   echo "</form>";

if (isset($check) & isset($sel_site)) {

    $q_check = "SELECT site_name, site_description, site_URL,
        country_name, city.city_ID, city.name, politics.politics
        from site
        left outer join country
            on site.country_id = country.country_id
        left outer join city
            on site.city_id = city.city_id
        left outer join politics
            on site.politics_id = politics.politics_id
        where site_ID = $sel_site";

    $r_check = mysql_db_query("newshound", $q_check);

    if ( $r_check & mysql_num_rows($r_check)>0 ) {
        $r = mysql_fetch_array($r_check);
        $site_URL = $r["site_URL"];  
        echo "<p>Profile for <a href=".$site_URL." target=new>$thissite</a>:  
        <p><table width = 50%>
            <tr>
                <td>Location</td>
                <td>Politics</td>
            </tr>
            <tr>
                <td>".$site_name."<br>
                    $country_name, "$city_name</td>
                <td>$politics</td>
            </tr>
        </table>
        $site_desc = ".$site_desc."<p>
    }
```
} else {
    echo "<p>Alert! No profile found for $thissite";
}

} # end if isset check and site

if ($r_c) {
    mysql_free_result($r_c);
}

?>

<hr><p><img src="smclearhound.gif">&nbsp;<a href="main.html">Back to main NewsHound page</a>
</body>
</html>
5. Where’s the party?
(http://dream.sims.berkeley.edu/newshound/holiday.php)
Source file holiday.php

<html>
<head>
<title>NewsHound Database Query - Where's the Party?</title>
<link rel=stylesheet type="text/css" href="styles.css">
</head>
<body>
<a name=top></a>
<h1><a href="main.html"><img src="clearhound.gif" border=0></a>&nbsp;NewsHound</h1><hr>
<h2>Where's the party?</h2>
<form action="$PHP_SELF" method=POST>
<?
  # the arguments to mysql_connect are hostname, 
  # username, and password
  mysql_connect("localhost", "newshound", "newshound");

  # get todays date in newshound format
  $dtarray = getdate(time());
  $today=$dtarray["year"].(01)$dtarray["mon"].(01)$dtarray["mday"];
  $show_today=$dtarray["mday"]." ".$dtarray["month"];  

  echo "<p>Pick a date: ";
  echo "<p><select name=pickday>
    
    if(isset($pickday)){
      $tmpday = strtotime( 
        substr($pickday,4,2).'/'.
        substr($pickday,6,2).'/'.
        substr($pickday,0,4) 
      );
      $showday = date("d F", $tmpday);
      echo "<option selected value=$pickday>$showday";
    } else {
      echo "<option selected value=$today>Today"
    }

    # starting date, number of days, index
    $s = mktime(0,0,0,1,1,2001);
    $max = 365;
    $i = 0;

    while ($i<$max) {
      $darray = getdate($s);
      $m = str_pad($darray["mon"],2,"0",STR_PAD_LEFT);
      $d = str_pad($darray["mday"],2,"0",STR_PAD_LEFT);
      $y = $darray["year"];  
      $currday=$y.$m.$d;
      $showday=$darray["mday"]." ".$darray["month"];  

      #echo "<p>DEBUG $currday DEBUG $showday";

      $s = strtotime($s + 86400);
      $i ++;
    }
  
<?

</form>
</body>
</html>
echo "<option value=$currday>$showday";

$s = mktime(0,0,0,$m,$d+1,$y);
$i++;
}

echo "</select>";
echo '<input type="submit" value="Woohoo">';

if(isset($pickday)){

$q_holiday = "SELECT country_name, holiday_name
FROM holiday, holiday_country, country
WHERE holiday.holiday_id = holiday_country.holiday_id
AND holiday_country.country_id = country.country_id
AND date=$pickday";

# the arguments to mysql_db_query are dbname, # and query string
$r_holiday = mysql_db_query("newshound", $q_holiday);

if (mysql_num_rows($r_holiday)) {
    echo "<p>";
    ## echo "DEBUG: $pickday <p>">
    echo "The party is in: <br>";
    echo "<ul>";
    while ($r = mysql_fetch_array($r_holiday)) {
        $country = $r["country_name"];
        $holiday = $r["holiday_name"];
        echo "<li>$country is celebrating $holiday</li>";
    }
    echo "</ul>";
    mysql_free_result($r_holiday);
} else {
    echo "<p>No party on this date. Throw your own!<p>";
    # echo "<p>No party today, $show_today.<p>";
}

} # end if isset pickday

?>
</form>

<hr><p><img src="smclearhound.gif">&nbsp;<a href="main.html">Back to main NewsHound page</a></p>
</body>
</html>
6. When's the party?
(http://dream.sims.berkeley.edu/newshound/country_holiday.php)
Source file country_holiday.php

```php
<?
mysql_connect("localhost", "newshound", "newshound");
if (isset($country_id) && $country_id!="") {
    # get currently selected country name
    $q_where = "SELECT country_name
           from country
           where country.country_id = "$country_id";"
    $r_where = mysql_db_query("newshound", $q_where);
    if (mysql_num_rows($r_where)>1) {
        echo "DEBUG error in query selecting name labels"
    }
    $r = mysql_fetch_array($r_where);
    $thiscountry = $r["country_name"]; $thiscountry_id = $country_id;
}

echo "<form action=$PHP_SELF method=POST>";
echo "<table>
<tr>
   <td align=right>
      Select Country: 
   </td><td>
   "Select name=country_id">

   $q_c = "SELECT country_id, country_name
          FROM country
          ORDER BY country_name";

   $r_c = mysql_db_query("newshound", $q_c);
   if ($r_c) {
       while ($r = mysql_fetch_array($r_c)) {
           $country_id = $r["country_id"];
           $country_name = $r["country_name"]; 
           if ($country_name==$thiscountry) {
               echo "<option selected value="$country_id">$country_name";
           } else {
               echo "<option value="$country_id">$country_name";
           }
       }
    }
```

else {
    echo "<option value="$country_id">$country_name"
}
}

} else {
    echo "</select><br>
$country_id = $thiscountry_id;

$country_id = $thiscountry_id;

if (isset($check)) {
    $q_check = "select distinct country_name, holiday.holiday_ID, holiday_name, monthname(date), year(date), dayofmonth(date) from holiday_country, country, holiday where holiday_country.country_ID = holiday.country_ID and holiday_country.country_ID = $country_id order by date";
    $r_check = mysql_db_query("newshound", $q_check);
    if ( $r_check && mysql_num_rows($r_check)>0 ) {
        echo "<p>Holidays celebrated in $thiscountry: <p>";
        echo "<p><table>\n<tr>\n<td>Holiday</td>\n<td>Date</td>\n</tr>
while ($r = mysql_fetch_array($r_check)) {

} else {
    echo "<option value="$country_id">$country_name"
}
} else {
    echo "</select><br>";

```php
if ($r_c) {
    mysql_free_result($r_c);
}

if ($r) {
    $holiday_name = $r['holiday_name'];
    $month = $r['monthname(date)'];
    $year = $r['year(date)'];
    $day = $r['dayofmonth(date)'];
    echo "<td>$holiday_name</td>";
    echo "<td>$month $day, $year</td></tr>";
}
```

Additional code for displaying holiday data in a table format.
Attachment D: MySQL Queries for data entry, update, and deletion

The following examples show typical queries used for adding, modifying, and deleting entries in the NewsHound database. These examples cover the tables for which data entry forms are not available.

### Table: politics

```
mysql> insert into politics set politics="green";
Query OK, 1 row affected (0.00 sec)

mysql> update politics set politics="blue" where politics="green";
Query OK, 1 row affected (0.00 sec)
Rows matched: 1  Changed: 1  Warnings: 0

mysql> delete from politics where politics="blue";
Query OK, 1 row affected (0.00 sec)
```

### Table: region

```
mysql> insert into region set region_name="oceania";
Query OK, 1 row affected (0.00 sec)

mysql> update region set region_name="eastasia" where region_name="oceania";
Query OK, 1 row affected (0.01 sec)
Rows matched: 1  Changed: 1  Warnings: 0

mysql> delete from region where region_name="eastasia";
Query OK, 1 row affected (0.00 sec)
```

### Table: site

```
mysql> insert into site (site_tag, site_name, site_description, site_URL, country_ID, city_ID, politics_ID) values ("sfbg", "The San Francisco Bay Guardian", "Free alternative weekly", "http://www.sfbg.com", 168, 186, 4);
Query OK, 1 row affected (0.00 sec)

mysql> update site set politics_ID = 2 where site_tag="sfbg";
Query OK, 1 row affected (0.00 sec)
Rows matched: 1  Changed: 1  Warnings: 0

mysql> delete from site where site_tag="sfbg";
Query OK, 1 row affected (0.00 sec)
```
Tables: leader and alt_leader

mysql> insert into leader (country_ID, date_updated, leader_name, leader_title, big_boss) values (1, 20011204, "Jimmy Durante", "Small Kahuna", '0');
Query OK, 1 row affected (0.00 sec)

mysql> insert into alt_leader (leader_ID, alt_name) values (183, "James Durante"神通);
Query OK, 1 row affected (0.00 sec)

mysql> update leader set country_ID=67 where leader_ID=183;
Query OK, 1 row affected (0.00 sec)
Rows matched: 1  Changed: 1  Warnings: 0

mysql> update alt_leader set alt_name="Jim-Bob Durante" where alt_name="James Durante"神通;
Query OK, 1 row affected (0.00 sec)
Rows matched: 1  Changed: 1  Warnings: 0

mysql> delete from alt_leader where leader_ID=183;
Query OK, 1 row affected (0.00 sec)

mysql> delete from leader where leader_ID=183;
Query OK, 1 row affected (0.00 sec)
Attachment E: MySQL Commands for Page Data Import

We have an on-going process of data collection and entry for page information (i.e., information captured from the set of sites on a daily basis). The data is collected and processed using a series of Perl scripts which use HTTP requests to fetch the pages, parse the HTML, and count total tokens, text word tokens, and image tags. An output file is produced in a format suitable for import into the database (comma delimited fields), as shown in the extract below:

```
abc, 20011201,17:30:00, abc.20011201, 3419, 88, 415
abc, 20011202,17:30:00, abc.20011202, 3198, 85, 439
allafrica, 20011201,17:30:00, allafrica.20011201, 4873, 189, 983
allafrica, 20011202,17:30:00, allafrica.20011202, 4883, 194, 954
arabnews, 20011201,17:30:00, arabnews.20011201, 1586, 3, 708
arabnews, 20011202,17:30:00, arabnews.20011202, 1768, 3, 871
asahi, 20011201,17:30:00, asahi.20011201, 3727, 226, 793
asahi, 20011202,17:30:00, asahi.20011202, 3736, 226, 796
bbc, 20011201,17:30:00, bbc.20011201, 3510, 190, 786
bbc, 20011202,17:30:00, bbc.20011202, 3570, 191, 835
```

Note that this file may contain one or more days worth of data.

**Import process steps**

First, any previous entries are cleared from the page_import table using the following MySQL command:

```
delete from page_import;
```

This text file is imported into the page_import table using the following MySQL command:

```
load data local infile 'pageimport.txt' into table page_import fields terminated by ',' (site_tag, date, time, full_text_file, token_count, image_count, word_count);
```

This contains all of the information needed for the page table except for the site_id; it instead contains the short version of the site name (site_tag). The following query obtains the site_id from the site table, combines it with the information from the page_import table, and places the results into the page table:

```
insert into page (site_id, date, time, word_count, full_text_file, image_count, token_count) select site_id, date, time, word_count, full_text_file, image_count, token_count from site, page_import where site.site_tag = page_import.site_tag;
```

Note that since the page_id field is defined as an auto_increment field, it is assigned automatically.
Attachment F: PHP Data Entry Forms

The following pages provide illustrations and source code for the following PHP-based forms:

1. Enter new country
   (http://dream.sims.berkeley.edu/newshound/add_country.php)

2. Enter or correct data for existing country
   (http://dream.sims.berkeley.edu/newshound/update_country.php)

3. View and enter alternate country names
   (http://dream.sims.berkeley.edu/newshound/add_alt_country.php)

4. Enter new city
   (http://dream.sims.berkeley.edu/newshound/add_city.php)

5. Enter or correct data for existing city
   (http://dream.sims.berkeley.edu/newshound/update_city.php)

6. View and enter alternate city names
   (http://dream.sims.berkeley.edu/newshound/add_alt_city.php)

7. Enter new holiday
   (http://dream.sims.berkeley.edu/newshound/add_holiday.php)

Queries are highlighted using a shaded box.
1. Enter new country
(http://dream.sims.berkeley.edu/newshound/add_country.php)
Source file add_country.php

```php
<?
    echo "<h1><a href=main.html><img src=clearhound.gif border=0></a>
    NewsHound</h1><hr>
    
    echo "<h2>Add a New Country</h2>
    
    mysql_connect("localhost", "newshound", "newshound");
    
    if (isset($dbentry)) {
        # check if required fields set before constructing query
        if ( (isset($country_name) && $country_name!="") &&
            (isset($region_id) && $region_id!="") ) {
            $q_insert = "INSERT into country SET ";
            $q_insert .= "country_name='$country_name', ";
            $q_insert .= "region_id='$region_id', ";
            
            # only add population if inputted
            if (isset($pop) && $pop!="") {
                $q_insert .= "pop='$pop' ";
                $continue = "yes";
            }
            # only add gdp if inputted
            if (isset($gdp) && $gdp!="") { 
                if ($continue="yes") {
                    $q_insert .= "-, ";
                }
                $q_insert .= "gdp='$gdp' ";
                # $continue = "yes";
            }
        }
        if ($q_insert) {
            # echo "DEBUG query is: $q_insert";
            $r_insert = mysql_db_query("newshound", $q_insert);
        }
        if ($r_insert) {
            $q_region = "SELECT region_name
            FROM region
            where region_id=$region_id";
        }
    }
?>

NewsHound Database Query - Add a New Country";
```

$r_region = mysql_db_query("newshound", $q_region);

$r = mysql_fetch_array($r_region);
    $region_name = $r["region_name"];

    echo "<p>You just added $country_name to $region_name with a population of $pop,000 and a GDP of $$gdp<br>";

    } # end if add entry

    echo "<form name='thisForm' action=$PHP_SELF method=POST>";

    echo "<table><tr><td align=right>
    echo "Enter Country Name: ";
    echo "</td><td>
    echo "<input type=text name='country_name' size=45>
    echo "</td></tr><tr>";

    echo "&nbsp;";
    echo "</td><td>
    echo "Enter Population (in thousands): 
    echo "</td><td>
    echo "&nbsp;";
    echo "</td>
    echo "</tr><tr>";

    echo "&nbsp;";
    echo "</td><td>
    echo "Enter Per Capita GDP (US$): 
    echo "</td><td>
    echo "&nbsp;";
    echo "</td>
    echo "</tr>";

    echo "&nbsp;";
    echo "</td><td>
    echo "<input type=text name='gdp' size=10>
    echo "</td></tr>";

    echo "</table>";

    } # end if add entry
```
$q_regions = "SELECT region_id, region_name
    FROM region
    ORDER BY region_name";

$r_regions = mysql_db_query("newshound", $q_regions);
if ($r_regions) {
    while ($r = mysql_fetch_array($r_regions)) {
        $region_id = $r["region_id"]; // Incorrect variable name
        $region_name = $r["region_name"]; // Incorrect variable name
        if ($region_name==$thisregion) {
            echo "<option selected
                value=$region_id>$region_name";
        } else {
            echo "<option value=$region_id>$region_name";
        }
    }
}
```

```
<?
<hr><p><img src="smclearhound.gif">&nbsp;<a href="main.html">Back to main NewsHound page</a>
</p>
</body>
</html>
```
2. Enter or correct data for existing country (http://dream.sims.berkeley.edu/newshound/update_country.php)
Source file update_country.php

```php
<html>
<head>
<title>NewsHound Database Query - Update Country Data</title>
<link rel=stylesheet type="text/css" href="styles.css">
<script type="text/javascript">
    function submitNow() {
        document.thisForm.submit();
    }
</script>
</head>
<body>
<h1><a href="main.html">NewsHound</a></h1>
<hr>
<?
    echo "<h2>Update Country Data</h2>";
    mysql_connect("localhost", "newshound", "newshound");

    if (isset($country_id) && $country_id!="") {
        # get currently selected country data
        $q_where = "SELECT country_name, country_id, country.region_id,  
                   region.region_name, gdp, pop  
                   from country, region  
                   where country.region_id = region.region_id  
                   and country_id = $country_id";

        $r_where = mysql_db_query("newshound", $q_where);
        if (mysql_num_rows($r_where)!=1) {
            echo "DEBUG error retrieving current data for id $country_id";
        }
        $r = mysql_fetch_array($r_where);

        $thiscountry = $r["country_name"];  
        $thiscountryid = $r["country_id"];  
        $thisregion = $r["region_name"];  
        $thisregionid = $r["region_id"];  
        $thisgdp = $r["gdp"];  
        $thispop = $r["pop"];  
        # echo "<br>DEBUG this is $thiscountry, $thisregion,  
        # regid = $thisregionid<br>";

        if (isset($dbentry)) {
            # check if any fields set before constructing query
            if ( (isset($country_name) && $country_name!="") ||  
                (isset($pop) && $pop!="" ) ||  
                (isset($gdp) && $gdp!="" ) ||  
                (isset($region_id) && $region_id=$thisregionid )
```

Attachment F
Page 7 of 34
$q_insert = "UPDATE country SET ";
if (isset($country_name) && !$country_name=="") {
    $q_insert .= "country_name='$country_name'";
    $continue = "yes";
}
# only insert region if actually modified
if (isset($region_id) && $region_id!=$thisregionid) {
    if ($continue=="yes") {
        $q_insert .= ", ";
    }
    $q_insert .= "region_id='$region_id' ";
    $continue = "yes";
}
if (isset($pop) && $pop!="") {
    if ($continue=="yes") {
        $q_insert .= ", ";
    }
    $q_insert .= "pop='$pop' ";
    $continue = "yes";
}
if (isset($gdp) && $gdp!="") {
    if ($continue=="yes") {
        $q_insert .= ", ";
    }
    $q_insert .= "gdp='$gdp' ";
    # $continue = "yes";
}
$q_insert .= " where country_id=$country_id";
} # end if any fields set
if ($q_insert) {
    # echo "DEBUG query is: $q_insert";
    $r_insert = mysql_db_query("newshound", $q_insert);
}
if ($r_insert) {
    $r_where = mysql_db_query("newshound", $q_where);
    if (mysql_num_rows($r_where)!=1) {
        echo "DEBUG error retrieving current data for id $country_id"
    }
    $r = mysql_fetch_array($r_where);
    $thiscountry = $r["country_name"];
    $thiscountryid = $r["country_id"];
    $thisregion = $r["region_name"];
    $thisregionid = $r["region_id"];
    $thisgdp = $r["gdp"];
    $thispop = $r["pop"];
    #echo "<br>DEBUG this is now $thiscountry, 
    #$thisregion, pop $pop, gdp $gdp<br>"
}
} # end if add entry
echo "<form name='thisForm' action=$PHP_SELF method=POST>">

echo "<table>
<tr>
<td align=right>

Select Country: 
</td><td>

<select name=country_id onChange='submitNow()'>

$q_countries = "SELECT country_id, country_name, country.region_id,  
region_name, gdp, pop 
FROM country, region  
where country.region_id=region.region_id  
ORDER BY country_name";

$r_countries = mysql_db_query("newshound", $q_countries);

if ($r_countries) {

### for debugging, do once instead of loop
### $r = mysql_fetch_array($r_countries);

while ($r = mysql_fetch_array($r_countries)) {
    $country_id = $r["country_id"];  
    if (!isset($thiscountryid)) {
        $thiscountryid=$country_id;
    }

    $country_name = $r["country_name"];

    # store all values for selected country
    # these are used for display below
    if ($country_id==$thiscountryid) {
        $thiscountry = $country_name;
        $thisgdp = $r["gdp"];  
        $thispop = $r["pop"];  
        $thisregion = $r["region_name"];  
        $thisregionid = $r["region_id"];  
        
        echo "<option selected value=$country_id>$country_name";
    }  
    else {
        echo "<option value=$country_id>$country_name";
    }
}
}

echo "</select><br>

echo "</td></tr>


echo "Enter Correct Country Name: ";
echo "</td><td>\n";
echo "<input type=text name='country_name' size=45>";
echo "</td><tr>\n";

echo "<tr><td>\n";
echo "&nbsp;";
echo "</td><tr>\n";

if ($thispop!="") {
    echo "Currently entered population is $thispop";
}
else {
    echo "No population information in database for $thiscountry";
}
echo "</td><tr>\n";

echo "<tr><td align=right>\n";
echo "Enter Correct Population (in thousands): ";
echo "</td><td>\n";
echo "<input type=text name='pop' size=10>";
echo "</td><tr>\n";

if ($thisgdp!="") {
    echo "Currently entered GDP is $thisgdp";
}
else {
    echo "No GDP information in database for $thiscountry";
}
echo "</td><tr>\n";

echo "<tr align=right>\n";
echo "Enter Correct Per Capita GDP (US$): ";
echo "</td><td>\n";
echo "<input type=text name='gdp' size=10>";
echo "</td><tr>\n";
```php
$w = "Select Correct Region: ";
$w = "$w";<td></td>
$w = "<select name=region_id>";
$w = "$w";
$w = "$w";
$w = "SELECT region_id, region_name 
FROM region 
ORDER BY region_name";

$r_regions = mysql_db_query("newshound", $q_regions);
if ($r_regions) {
    while ($r = mysql_fetch_array($r_regions)) {
        $region_id = $r["region_id"];  
        $region_name = $r["region_name"];  
        if ($region_name==$thisregion) {
            $w = "$w";
        } else {
            $w = "$w";
        }
    }
}
$w = "$w";
$w = "&nbsp;";
$w = "&nbsp;";
$w = "&nbsp;";
$w = "<input type=submit name='dbentry' value='Submit Change'>";
$w = "</form>";
if ($r_countries) {
    mysql_free_result($r_countries);
}
?>
<br><p><img src="smclearhound.gif">&nbsp;<a href="main.html">Back to main NewsHound page</a>
</body>
</html>
```
3. View and enter alternate country names
   (http://dream.sims.berkeley.edu/newshound/add_alt_country.php)
Source file add_alt_country.php

```php
<?
echo "<h2>Add an entry: Alternate country name</h2>";
mysql_connect("localhost", "newshound", "newshound");
if (isset($_POST['submit']) && isset($_POST['alt_name'])) {
    $q_insert = "INSERT INTO alt_country VALUES ('$_POST[country_id]', '" . $_POST['alt_name'] . ")";
    $r_insert = mysql_db_query("newshound", $q_insert);
    if ($r_insert) {
        echo "$alt_name was added as an alternate name for $thiscountry";
    }
} else if (isset($_POST['check'])) {
    $q_check = "select country_name, alt_name
        from country, alt_country
        where country.country_id = alt_country.country_id
        and alt_country.country_id = $_POST[country_id]";
    $r_check = mysql_db_query("newshound", $q_check);
    if ( ($r_check && mysql_num_rows($r_check)>0 ) {
```

```php
```
echo "<p>Found the following alternate name(s) for $thiscountry: <ul>";
while ($r = mysql_fetch_array($r_check)) {
    $alt_name = $r["alt_name"];
    echo "<li>$alt_name ";
}
echo "</ul>";
} else {
echo "<p>No alternate names found for $thiscountry";
}
} # end if check current entries
} # end if isset country id

echo "<form action=$PHP_SELF method=POST">";

echo "<table>
<tr>
<td align=right><br>
Select Country: </td><td>
<select name=country_id>
$q_c = "SELECT country_id, country_name
FROM country
ORDER BY country_name";

$r_c = mysql_db_query("newshound", $q_c);
if ($r_c) {
    while ($r = mysql_fetch_array($r_c)) {
        $country_id = $r["country_id"];
        $country_name = $r["country_name"];
        if ($country_name==$thiscountry) {
            echo "<option selected value=$country_id>$country_name";
        } else {
            echo "<option value=$country_id>$country_name";
        }
    }
} echo "</select><br>";
echo "</td></tr>
<tr><td><br>
Enter Alternate Name: </td><td>
<input type=text name='alt_name' size=45>
</td></tr>
</table>
";
</form>";
```
echo "<tr><td>
 echo "&nbsp;"
 echo "</td><td>
 echo "<input type=submit name='submit' value='Add entry'>";
 echo "</td><tr</td>
";

echo "</table>
";
echo "</form>
";

if ($r_c) {
   mysql_free_result($r_c);
}

?>
</hr><p><img src="smclearhound.gif">&nbsp;<a href="main.html">Back to main NewsHound page</a></p>
</body>
</html>
```
4. **Enter new city**
   (http://dream.sims.berkeley.edu/newshound/add_city.php)
Source file add_city.php

```php
<html>
<head>
<title>NewsHound Database Query - Add a New City</title>
<link rel=stylesheet type="text/css" href="styles.css">
</head>

<body>
<?
    echo "<h1><a href=main.html><img src=clearhound.gif border=0></a>
NewsHound</h1><hr>
    echo "<h2>Add a New City</h2>";
mysql_connect("localhost", "newshound", "newshound");

    if (isset($submit)) {
        if ( (isset($city_name) && $city_name!="") &&
            (isset($cap_city) && $cap_city!="") &&
            (isset($country_id) && $country_id!="")) {
            # echo "DEBUG inserting query";

            $q_insert = "INSERT into city
                SET city_name='$city_name',
                cap_city='$cap_city',
                country_id='$country_id'";

        }
    }

    if ($q_insert) {
        $r_insert = mysql_db_query("newshound", $q_insert);
    }

    if ($r_insert) {
        $q_country = "SELECT country_name
                FROM country
                where country_id=$country_id";

        $r_country = mysql_db_query("newshound", $q_country);
        $r = mysql_fetch_array($r_country);
        $country_name = $r["country_name"];  

        if ($cap_city == '1') {
            echo "<p>$city_name has been added to $country_name. It is marked as the capital of $country_name.</p>";
        } else {
            echo "<p>$city_name has been added to $country_name. It is not the capital of $country_name.</p>";
        }
    }

} # end if add entry
```
echo "<form action=$PHP_SELF method=POST>">

echo "<table>
<tr>
<td align=right>
"Enter New City Name: ";
</td><td>
<input type=text name='city_name' size=45>
</td></tr>

<tr><td>
 </td><td>
 </td>
</tr>
<tr><td align=right>
"This is a capital city."
</td><td>
<input type=radio name='cap_city' value='1'>
</td>
</tr>
<tr><td>
"This is not a capital city."
</td><td>
<input type=radio name='cap_city' value='0' checked>
</td>
</tr>
<tr><td>
 </td><td>
 </td>
</tr>
<tr><td align=right>
"Select Country: ";
</td><td>
<select name=country_id>

$q_country = "SELECT country_id, country_name
FROM country
ORDER BY country_name";

$r_country = mysql_db_query("newshound", $q_country);

if ($r_country) {
    while ($r = mysql_fetch_array($r_country)) {
        $country_id = $r["country_id"];  
        $country_name = $r["country_name"];  
        if ($country_name==$thiscountry) {
            echo "<option selected value=$country_id>$country_name";
        } else {
            echo "<option value=$country_id>$country_name";
        }
    }
}
[\n\n]}\n\n\n}\n\n\n\n\n\n\n\n\n\nif ($r_cities) {\n    mysql_free_result($r_cities);\n}\n\n<?\n
<hr><p><img src="smclearhound.gif"> &nbsp; <a href="main.html">Back to main NewsHound page</a></p>

</body>

</html>
5. Enter or correct data for existing city
(http://dream.sims.berkeley.edu/newshound/update_city.php)
Source file update_city.php

```php
<html>
<head>
<title>NewsHound Database Query - Update City Data</title>
<link rel=stylesheet type="text/css" href="styles.css">
</head>

<body>
<h1><a href="main.html"><img src="clearhound.gif" border=0></a>&nbsp;NewsHound</h1><hr>

<?

// get city, country names to use in text output
$q_where = "SELECT city_name, country_name, cap_city
from city, country
where city.country_id = country.country_id
and city.city_id = $city_id";

$r_where = mysql_db_query("newshound", $q_where);
if (mysql_num_rows($r_where)>1) {
    echo "DEBUG error in query selecting name labels";
}
$r = mysql_fetch_array($r_where);
$thiscity = $r["city_name"];
$thiscountry = $r["country_name"];
$thiscapCityFlag = $r["cap_city"];
if (isset($submit)) {
    if (isset($city_name) && !($city_name=="")) {
        if (isset($cap_city)) {
            $q_insert = "UPDATE city
set city_name='$city_name',
cap_city='$cap_city'";
            where city_id=$city_id";
        } else {
            $q_insert = "UPDATE city
set city_name='$city_name'
where city_id=$city_id";
        }
    } else {
        $q_insert = "UPDATE city
set city_name='$city_name'
where city_id=$city_id";
    }
```
else if (isset($cap_city)) {
    $q_insert = "UPDATE city
        SET cap_city='$cap_city'
        where city_id=$city_id";
}

if ($q_insert) {
    $r_insert = mysql_db_query("newshound", $q_insert);
}
$r_where = mysql_db_query("newshound", $q_where);
if (mysql_num_rows($r_where)>1) {
    echo "DEBUG error in query selecting name labels";
}
$r = mysql_fetch_array($r_where);
$thiscity = $r["city_name"];
$thiscountry = $r["country_name"];
$thiscapCityFlag = $r["cap_city"];
if ($r_insert) {
    if ($thiscapCityFlag == '1') {
        echo "<p>City name has been updated to
            $thiscity, $thiscountry.
            It is marked as the capital of
            $thiscountry.</p>";
    } else {
        echo "<p>City name has been updated to
            $thiscity, $thiscountry.</p>";
    }
}

} # end if add entry
else if (isset($check)) {

    $q_check = "select city_name, alt_name, cap_city
        from city, alt_city
        where city.city_id = alt_city.city_id
        and alt_city.city_id = $city_id";
    $r_check = mysql_db_query("newshound", $q_check);
    if ( $r_check && mysql_num_rows($r_check)>0 ) {
        echo "<p>Found the following alternate name(s) for
            $thiscity, $thiscountry: <ul>";
        while ($r = mysql_fetch_array($r_check)) {
            $city_name = $r["city_name"];
            $alt_name = $r["alt_name"];
            echo "<li>$alt_name ";
        }
        echo "\n</ul>";
    }
else {
        echo "<p>No alternate city names found for $thiscity, $thiscountry";
    } # end if city_id
}

echo "<form action=$PHP_SELF method=POST>";

echo "<table>n<tr>n<td align=right>n";
echo "Select City: ";
echo "</td><td>";
echo "<select name=city_id>";

$q_cities = "SELECT city_id, city_name, country_name, cap_city
FROM city, country
where city.country_id = country.country_id
ORDER BY city_name";

$r_cities = mysql_db_query("newshound", $q_cities);

if ($r_cities) {
    while ($r = mysql_fetch_array($r_cities)) {
        $city_id = $r["city_id"];
        $city_name = $r["city_name"];
        $country_name = $r["country_name"];
        $capCityFlag = $r["cap_city"];

        if ($capCityFlag == '1') {
            if ($city_name==$thiscity) {
                echo "<option selected value=$city_id>$city_name,
                 $country_name (capital)";
            }
            else {
                echo "<option value=$city_id>$city_name,
                 $country_name (capital)";
            }
        }
        else {
            if ($city_name==$thiscity) {
                echo "<option selected value=$city_id>$city_name,
                 $country_name";
            }
            else {
                echo "<option value=$city_id>$city_name,
                 $country_name";
            }
        }
    }
}
} # end if $r_cities


```php
)

echo "</select><br>");
echo "</td></tr><br>

echo "<tr><td>
echo "&nbsp;";
echo "</td><tr><td>
echo "&nbsp;";
echo "</tr><td><tr>
echo "&nbsp;";
echo "</td><tr>
echo "&nbsp;";
echo "</td><tr>

```

```php
echo "</td><tr><td align=right>
echo "Enter Correct City Name: ";
echo "</td><tr><td>
echo "<input type=text name='city_name' size=45>"
echo "</td></tr><br>

echo "&nbsp;";
echo "</td><tr><td>
echo "&nbsp;";
echo "</td><tr><td>
echo "&nbsp;";
echo "</td><tr>

```

```php
echo "&nbsp;";
echo "<td align=right>
echo "Yes, this is a capital city.";
echo "</td><tr><td>
echo "<input type=radio name='cap_city' value='1'>"
echo "</td></tr><br>

```

```php
echo "&nbsp;";
echo "<td align=right>
echo "No, this is not a capital city.";
echo "</td><tr><td>
echo "<input type=radio name='cap_city' value='0'>"
echo "</td></tr><br>

```

```php
echo "<tr><td>
echo "&nbsp;";
echo "</td><tr><td>
```
echo "\n\n\n\n$\n\n\n\n$r_cities) {
    mysql_free_result($r_cities);
}"
6. View and enter alternate city names
   (http://dream.sims.berkeley.edu/newshound/add_alt_city.php)
Source file add_alt_city.php

```php
<?
    echo "<h2>Add an entry: Alternate city name</h2>";
    mysql_connect("localhost", "newshound", "newshound");
    if (isset($city_id)) {
        # get city, country names to use in text output
        $q_where = "SELECT city_name, country_name
                    from city, country
                    where city.country_id = country.country_id
                    and city.city_id = $city_id";

        $r_where = mysql_db_query("newshound", $q_where);
        if (mysql_num_rows($r_where)>1) {
            echo "DEBUG error in query selecting name labels";
        }
        $r = mysql_fetch_array($r_where);
        $thiscity = $r["city_name"];
        $thiscountry = $r["country_name"];
        if (isset($submit) && isset($alt_name)) {
            $q_insert = "INSERT INTO alt_city VALUES ('$city_id', '$alt_name')";
            $r_insert = mysql_db_query("newshound", $q_insert);
            if ($r_insert) {
                echo "<p>$alt_name was added as an alternate name for $thiscity, $thiscountry</p>";
            }
        } # end if add entry
    } else if (isset($check)) {
        $q_check = "select city_name, alt_name
                    from city, alt_city
                    where city.city_id = alt_city.city_id
                    and alt_city.city_id = $city_id";
        $r_check = mysql_db_query("newshound", $q_check);
    }
?>
```

if ( $r_check && mysql_num_rows($r_check)>0 ) {
    echo "<p>Found the following alternate name(s) for $thiscity, $thiscountry: <ul>
    while ($r = mysql_fetch_array($r_check)) {
        $city_name = $r["city_name"];
        $alt_name = $r["alt_name"];
        echo "<li>$alt_name ";
    }
    echo "\n</ul>";
} else {
    echo "<p>No alternate city names found for $thiscity, $thiscountry";
}
} # end if city_id

$g_cities = "SELECT city_id, city_name, country_name
    FROM city, country
    where city.country_id = country.country_id
    ORDER BY city_name";

$r_cities = mysql_db_query("newshound", $q_cities);
if ($r_cities) {
    while ($r = mysql_fetch_array($r_cities)) {
        $city_id = $r["city_id"];
        $city_name = $r["city_name"];
        $country_name = $r["country_name"];
        if ($city_name==$thiscity) {
            echo "<option selected value=$city_id>$city_name, $country_name";
        } else {
            echo "<option value=$city_id>$city_name, $country_name";
        }
    }
    echo "</select><br>
    echo "</td></tr><n";

    echo "<tr><td>
    echo " "
    echo "</td>
    echo "</tr><n";

    echo "<tr><td>
    echo "&nbsp;";
    echo "</td><td>

echo "<input type=submit name='check' value='Check existing entries'>";
echo "</td><n</tr><n"

echo "<tr><td align=right><n"
echo "Enter Alternate Name: 
echo "</td><n"
echo "<input type=text name='alt_name' size=45>
echo "</td></tr<n"

echo "</tr><n"
echo "&nbsp;"
echo "</td><n"
echo "<input type=submit name='submit' value='Add entry'>"
echo "</td><n</tr><n"

echo "</table>n"
echo "</form">

if ($r_cities) {
    mysql_free_result($r_cities);
}

?>

<hr><p><img src="smclearhound.gif">&nbsp;<a href="main.html">Back to main NewsHound page</a>

</p>
</body>
</html>
7. Enter new holiday (http://dream.sims.berkeley.edu/newshound/add_holiday.php)
Source file add_holiday.php

```php
<?
  echo "<h2>Add an entry: Holiday data</h2>";
  mysql_connect("localhost", "newshound", "newshound");

  if (isset($date) && isset($country_id)) {
    # get selected country name for display
    $q_where = "SELECT country_name
                FROM country
                WHERE $country_id = country.country_id";

    $r_where = mysql_db_query("newshound", $q_where);
    if (mysql_num_rows($r_where)>1) {
      echo "DEBUG error in query selecting name labels"
    }
    $r = mysql_fetch_array($r_where);
    $thiscountry = $r["country_name"];  

    $seldate = mktime(0,0,0,substr($date,4,2),
                      substr($date,6,2),substr($date,0,4));
    $seldarray = getdate($seldate);
    $thisdate=$seldarray["mday"] . ".".$seldarray["month"] . "." .
               $seldarray["year"];  
    # echo "<br>DEBUG selected date is $date AKA $thisdate<br>";

    if (isset($submit) && isset($holiday_name) && $holiday_name!="")
      {
        $q_in1 = "INSERT INTO holiday (date, holiday_name)
                 VALUES ($date, "$holiday_name")";

        # echo "DEBUG q_in1 is $q_in1<br>"
        $r_in1 = mysql_db_query("newshound", $q_in1);
        if ($r_in1) {

          $q_in2 = "INSERT INTO holiday_country
                   VALUES ($country_id, last_insert_id())";

          # echo "DEBUG q_in2 is $q_in2<br>"
          $r_in2 = mysql_db_query("newshound", $q_in2);

        }
      }
  }
?>
```
echo "<p>$holiday_name was added as a holiday on $thisdate in $thiscountry</p>";
}
}
# end if add entry
else if (isset($check_date)) {

$q_check_date = "select holiday_name, country_name
from holiday, holiday_country, country
where holiday.holiday_id = holiday_country.holiday_id
and holiday_country.country_id = country.country_id
and holiday.date = $date";

$r_check_date = mysql_db_query("newshound", $q_check_date);

if ( $r_check_date && mysql_num_rows($r_check_date)>0 ) {
  echo "<p>Found the following holidays on $thisdate: <ul>"
  while ($r = mysql_fetch_array($r_check_date)) {
    $holiday_name = $r["holiday_name"];
    $country_name = $r["country_name"];
    # $dispdate = $r["dispdate"];
    echo "<li>$holiday_name in $country_name ";
  }
  echo "\n</ul>";
} 
else {
  echo "<p>No holidays found on $thisdate";
}
}
else if (isset($check_country)) {

$q_check_c = "select holiday_name, date,
DATE_FORMAT(date,'%d %M %Y') AS dispdate,
country_name
from holiday, holiday_country, country
where holiday.holiday_id = holiday_country.holiday_id
and holiday_country.country_id = country.country_id
and holiday_country.country_id = $country_id";

# echo "<br>DEBUG query is $q_check_c<br>";
$r_check_c = mysql_db_query("newshound", $q_check_c);

if ( $r_check_c && mysql_num_rows($r_check_c)>0 ) {
  echo "<p>Found the following holidays in $thiscountry: <ul>"
  while ($r = mysql_fetch_array($r_check_c)) {
    $holiday_name = $r["holiday_name"];
    $date = $r["date"]; 
    $dispdate = $r["dispdate"]; 
    echo "<li>$holiday_name on $dispdate ";
  }
  echo "\n</ul>";
} 
else {
  echo "<p>No holidays found in $thiscountry";
}
echo "<form action=$PHP_SELF method=POST>
";

echo "\n<table><tr><td align=right><n">

echo "Select Date: ";
echo "</td></tr>

# starting date, number of days, index
$s = mktime(0,0,0,1,1,2001);
$max = 365;
i = 0;

while ($i<$max) {
    $darray = getdate($s);
    $m = str_pad($darray["mon"],2,"0",STR_PAD_LEFT);
    $d = str_pad($darray["mday"],2,"0",STR_PAD_LEFT);
    $y = $darray["year"];  
    $currday=$y.$m.$d;
    # $showday=$darray["mday"] . " ". $darray["month"] .  " ". $darray["year"];
    $showday=$darray["mday"] .  " ". $darray["month"] .  " ". $darray["year"];
    if ($showday==$thisdate) {
        echo "<option selected value=$currday>$showday";
    } else {
        echo "<option value=$currday>$showday";
    }
    $s = mktime(0,0,0,$m,$d+1,$y);
    $i++;
}

echo "</select><br>
";

while ($i<$max) {
    $darray = getdate($s);
    $m = str_pad($darray["mon"],2,"0",STR_PAD_LEFT);
    $d = str_pad($darray["mday"],2,"0",STR_PAD_LEFT);
    $y = $darray["year"];
    $currday=$y.$m.$d;
    # $showday=$darray["mday"] .  " ". $darray["month"] .  " ". $darray["year"];
    $showday=$darray["mday"] .  " ". $darray["month"] .  " ". $darray["year"];
    if ($showday==$thisdate) {
        echo "<option selected value=$currday>$showday";
    } else {
        echo "<option value=$currday>$showday";
    }
    $s = mktime(0,0,0,$m,$d+1,$y);
    $i++;
}

$q_c = "SELECT country_name, country_id
FROM country
ORDER BY country_name";

$r_c = mysql_db_query("newshound", $q_c);"
if ($r_c) {
    while ($r = mysql_fetch_array($r_c)) {
        $country_id = $r["country_id"];
        $country_name = $r["country_name"];
        if ($country_name==$thiscountry) {
            echo "<option selected value=$country_id>$country_name";
        } else {
            echo "<option value=$country_id>$country_name";
        }
    }
}

echo "</select><br>";
echo "</td></tr>

echo "<tr><td align=right><input type=submit name='check_country'
value='Check holidays in this country'>";

if ($r_cities) {
    mysql_free_result($r_cities);
} else {
    mysql_free_result($r_cities);
}

</table>
</form>
Attachment G: PHP Reports

The following pages provide illustrations and source code for the following PHP-based reports:

1. Countries by region
   (http://dream.sims.berkeley.edu/newshound/reg_country.php)

2. Capital city by country
   (http://dream.sims.berkeley.edu/newshound/country_city.php)

3. Leaders by country
   (http://dream.sims.berkeley.edu/newshound/country_leader.php)

4. Sites by region
   (http://dream.sims.berkeley.edu/newshound/site_reg.php)

5. Sites by political affiliation
   (http://dream.sims.berkeley.edu/newshound/site_politics.php)

Queries are highlighted using a shaded box.
1. Countries by Region
(http://dream.sims.berkeley.edu/newshound/reg_country.php)
Middle East / North Africa

- Afghanistan
- Bahrain
- Cyprus
- Egypt
- Iran
- Iraq
- Israel
- Jordan
- Kuwait
- Lebanon
- Libya
- Morocco
- Oman
- Qatar
- Saudi Arabia
- Syria
- United Arab Emirates
- Yemen

North America

- Canada
- U.S.A.

Back to main NewsHound page
Source file reg_country.php

```php
<?
    # the arguments to mysql_connect are hostname, 
    # username, and password
    mysql_connect("localhost", "newshound", "newshound");
    
    # query that obtains all regions containing at least one country 
    # (i.e., skip dummy region "World"); 
    # this is used for region links at top of page
    $q_reg_list = "SELECT count(country.country_ID), region.region_name 
    FROM region, country 
    where region.region_ID = country.region_ID 
    GROUP BY region.region_ID 
    ORDER BY region_name";
    
    # query that obtains set of region, country names 
    # main data used in page
    $q_country = "SELECT DISTINCT region_name, country_name FROM region, country 
    WHERE region.region_ID = country.region_ID 
    ORDER BY region_name, country_name";
    
    # the arguments to mysql_db_query are dbname, 
    # and query string
    $r_reg_list = mysql_db_query("newshound", $q_reg_list);
    $r_country = mysql_db_query("newshound", $q_country);
    
    if ($r_country) {
        if ($r_reg_list) {
            echo "<p>";
            while ($r = mysql_fetch_array($r_reg_list)) {
                $region_name = $r["region_name"];   
                echo "<a href="#$region_name">$region_name</a> 
                &nbsp;&nbsp;&nbsp;";
            }
            mysql_free_result($r_reg_list);
        }
        echo "<p>";
    }
    
    $curr_region="";
    $first="yes";
```
while ($r = mysql_fetch_array($r_country)) {
    $region_name = $r['region_name'];
    $country_name = $r['country_name'];
    if ($curr_region!=$region_name) {
        echo "</ul>";
        if ($first=="no") {
            echo "<p><a href=#top>Return to top</a>";
        }
        else { $first="no"; }
        echo "<a name="$region_name"></a>";
        echo "<h3>$region_name</h3><ul>";
        $curr_region=$region_name;
        echo "<li>$country_name";
    }
    echo "</li>";
    mysql_free_result($r_country);
} else {
    echo "No data.";
}
</a>
<a href=#top>Return to top</a>
<p></p><hr><p><img src="smclearhound.gif">&nbsp;<a href="main.html">Back to main NewsHound page</a></p>
2. Capital City by Country
(http://dream.sims.berkeley.edu/newshound/country_city.php)
Source file country_city.php

<html>
<head>
<title>NewsHound Database Query - Cities by Country</title>
<link rel=stylesheet type="text/css" href="styles.css">
</head>

<body>
<a name=top></a>
<h1><a href="main.html"><img src="clearhound.gif" border=0></a>&nbsp;NewsHound</h1><hr>
<h2>Capital Cities by Country</h2>
<?
# the arguments to mysql_connect are hostname,
# username, and password
mysql_connect("localhost", "newshound", "newshound");

# query that obtains all letters that start at least one country name
# this is used for links at top of page
$q_reg_list = "SELECT distinct substring(country_name,1,1)
as letter from country";

# query that obtains set of country, city names for cap cities
# main data used in page
$q_country = "SELECT country_name, city_name, cap_city FROM country, city
WHERE country.country_ID = city.country_ID and cap_city like 'l'
ORDER BY country_name, city_name";

# the arguments to mysql_db_query are dbname,
# and query string
$r_reg_list = mysql_db_query("newshound", $q_reg_list);
$r_country = mysql_db_query("newshound", $q_country);

if ($r_country) {
  if ($r_reg_list) {
    echo "<p>
    while ($r = mysql_fetch_array($r_reg_list)) {
      $letter = $r["letter"];     
      echo "<a href="#$letter'>$letter</a>
    }
    mysql_free_result($r_reg_list);
  }
  echo "<p>";

  $curr_letter="";
  $first="yes";
  while ($r = mysql_fetch_array($r_country)) {
    $country_name = $r["country_name"];UTF-8 encoding
    $city_name = $r["city_name"];UTF-8 encoding
    $cap_city = $r["cap_city"];UTF-8 encoding

  }
$letter=substr($country_name,0,1);
if ($curr_letter!=$letter){
  echo "</ul>");
  if ($first=="no") {
    echo "<p><a href=#top>Return to top</a>");
  }
  else { $first="no"; }
  echo "<a name=$letter></a>");
  echo "<h3>$letter</h3><ul>");
  $curr_letter=$letter;
}
  echo "<li>$country_name : $city_name";}
  echo "</ul>");
  mysql_free_result($r_country);
} else {
  echo "No data.";
}
}?

<a href=#top>Return to top</a>

<p></p><hr><p><img src="smclearhound.gif">&nbsp;<a href="main.html">Back to main NewsHound page</a>
</p>
<body>
</html>
3. Leaders by Country
(http://dream.sims.berkeley.edu/newshound/country_leader.php)
<html>
<head>
<title>NewsHound Database Query - Leaders by Country</title>
<link rel="stylesheet type="text/css" href="styles.css"/>
</head>

<body>
<a name=top></a>
<h1><a href="main.html"><img src="clearhound.gif" border=0></a>&nbsp;NewsHound</h1><hr>
<h2>Primary Leaders by Country</h2>

<?

# the arguments to mysql_connect are hostname, 
# username, and password
mysql_connect("localhost", "newshound", "newshound");

# query that obtains all letters that start at least one country name 
# this is used for links at top of page
$q_reg_list = "SELECT distinct substring(country_name,1,1) 
as letter from country";

# query that obtains set of leader, region, country names 
# main data used in page
$q_country = "SELECT DISTINCT region_name, country_name, 
leader_name, leader_title, big_boss 
FROM region, country, leader 
WHERE region.region_ID = country.region_ID 
and country.country_ID = leader.country_ID 
and leader.big_boss like '1' 
ORDER BY country_name";

# the arguments to mysql_db_query are dbname, 
# and query string
$r_reg_list = mysql_db_query("newshound", $q_reg_list);
$r_country = mysql_db_query("newshound", $q_country);

if ($r_country) {
    if ($r_reg_list) {
        echo "<p>";
        while ($r = mysql_fetch_array($r_reg_list)) {
            $letter = $r["letter"]; 
            echo "<a href=#$letter>$letter</a> 
            &nbsp;&nbsp;&nbsp;&nbsp;";
        }
        mysql_free_result($r_reg_list);
    }
    echo "&lt;p&gt;";
    $curr_letter="";
    $first="yes";
    while ($r = mysql_fetch_array($r_country)) {

Attachment G
Page 10 of 17
$country_name = $r['country_name'];
$leader_name = $r['leader_name'];
$leader_title = $r['leader_title'];
$letter=substr($country_name,0,1);
if ($curr_letter!=$letter){
echo "</ul>";
   if ($first=="no") {
      echo "<p><a href=#top>Return to top</a>";
   } else { $first="no"; }
   echo "<a name=$letter></a>";
   echo "<h3>$letter</h3><ul>";
   $curr_letter=$letter;
}
echo "<li>$country_name : $leader_title $leader_name";
}
echo "</ul>";
mysql_free_result($r_country);
} else {
   echo "No data.";
}
?>

<a href=#top>Return to top</a>

<p></p><hr><p><img src="smclearhound.gif">&nbsp;<a href="main.html">Back to main NewsHound page</a>
</p></body>
</html>
4. Sites by Region
(http://dream.sims.berkeley.edu/newshound/site_reg.php)
Source file site_reg.php

```php
<html>
<head>
    <title>NewsHound Database Query - Sites by Region</title>
    <link rel=stylesheet type="text/css" href="styles.css">
</head>

<body bgcolor="#ffffff">
<a name=top></a>
<h1><a href="main.html"><img src="clearhound.gif" border=0></a>&nbsp;NewsHound</h1>
<hr>
<h2>Sites by Region</h2>

<?
    # the arguments to mysql_connect are hostname, 
    # username, and password
    mysql_connect("localhost", "newshound", "newshound");

    # query that obtains all regions containing at least one country 
    # (i.e., skip dummy region "World"); 
    # this is used for region links at top of page
    $q_reg_list = "SELECT count(country.country_ID), region.region_name 
        FROM region, country 
        where region.region_ID = country.region_ID 
        GROUP BY region.region_ID 
        ORDER BY region_name";

    # query that obtains set of sites, regions, country names 
    # main data used in page
    $q_sites = "select site_name, site_URL, region_name, country_name 
        from site left outer join country 
        ON site.country_id=country.country_id 
        left outer join region 
        on country.region_id=region.region_id 
        ORDER BY region_name, country_name, site_name";

    # the arguments to mysql_db_query are dbname, 
    # and query string
    $r_reg_list = mysql_db_query("newshound", $q_reg_list);
    $r_sites = mysql_db_query("newshound", $q_sites);

    if ($r_sites) {
        if ($r_reg_list) {
            echo "<p>
            while ($r = mysql_fetch_array($r_reg_list)) {
                $region_name = $r["region_name"];
                echo "<a href="#$region_name"$region_name</a>
            &nbsp;&nbsp;&nbsp;&nbsp;";
            }
            mysql_free_result($r_reg_list);
        }
        echo "<br><br>";
    }
</body>
</html>
```
$curr_region="";
$first="yes";
while ($r = mysql_fetch_array($r_sites)) {
    $site_name = $r["site_name"];
    $site_url = $r["site_URL"];
    $region_name = $r["region_name"];

    # if no region, site is international
    if ($region_name==""") {
        $region_name="International";
    }

    $country_name = $r["country_name"];
    if ($curr_region!=$region_name){
        echo "</ul>";
        if ($first=="no") {
            echo "<p><a href=#top>Return to top</a><br><br>";
        }
        else { $first="no"; }
        echo "<a name="$region_name"></a>";
        echo "<h3>$region_name</h3><ul>";
        $curr_region=$region_name;
    }
    echo "<li>";
    echo "<a href="$site_url" target=new>";
    if ($region_name=="International"){
        echo "$site_name";
    } else {
        echo "$site_name ($country_name)";
    }
    echo "</a>";
}
echo "</ul>";
mysql_free_result($r_sites);
} else {
    echo "No data.";
}

<a href=#top>Return to top</a>

<hr><p><img src="smclearhound.gif">&nbsp;<a href="main.html">Back to main NewsHound page</a></p>
</body>
</html>
5. Sites by political affiliation
(http://dream.sims.berkeley.edu/newshound/site_politics.php)
<html>
<head>
  <title>NewsHound Database Query - Sites by Politics</title>
  <link rel=stylesheet type="text/css" href="styles.css">
</head>

<body bgcolor=#ffffff>
  <a name=top></a>
  <h1><a href="main.html"><img src="clearhound.gif" border=0></a>&nbsp;NewsHound</h1><hr>
  <h2>Sites by Politics</h2>

  <?
  # the arguments to mysql_connect are hostname, username, and password
  mysql_connect("localhost", "newshound", "newshound");

  # query that obtains all political affiliations containing at least one country
  # this is used for politics links at top of page
  $q_pol_list = "SELECT count(site.site_ID), politics.politics
  FROM site, politics
  WHERE site.politics_ID = politics.politics_ID
  GROUP BY politics.politics_ID
  ORDER BY politics.politics";

  # query that obtains set of site names, politics
  # main data used in page
  $q_sites = "select country_name, site_name, site_URL, politics
  from site, country left outer join politics
  on politics.politics_id=site.politics_id
  where site.country_ID = country.country_ID
  ORDER BY politics, site_name";

  # the arguments to mysql_db_query are dbname, # and query string
  $r_pol_list = mysql_db_query("newshound", $q_pol_list);
  $r_sites = mysql_db_query("newshound", $q_sites);

  if ($r_sites) {
    if ($r_pol_list) {
      echo "<p>";
      while ($r = mysql_fetch_array($r_pol_list)) {
        $politics = $r["politics"];  
        echo "<a href="#$politics">$politics</a>
        &nbsp;&nbsp;&nbsp;";
      }
      mysql_free_result($r_pol_list);
    }
    echo "<br><br>";
  }

  $curr_pol="";
  $first="yes";

</body>
</html>
while ($r = mysql_fetch_array($r_sites)) {
    $site_name = $r["site_name"]; 
    $site_url = $r["site_URL"]; 
    $politics = $r["politics"]; 
    $country_name = $r["country_name"]; 
    if ($politics) {
        if ($curr_pol != $politics) {
            echo "</ul>";
            if ($first == "no") {
                echo "<p><a href=#top>Return to top</a><br><br>");
            } else {
                $first = "no";
            }
            echo "<a name="politics"></a>"; 
            echo "<h3>$politics</h3><ul>"; 
            $curr_pol = $politics;
        }
        echo "<li>"; 
        echo "<a href="$site_url target=new">"; 
        echo "$site_name ($country_name)"; 
        echo "</a>"; 
        echo "</li>";
    }
    mysql_free_result($r_sites);
} else {
    echo "No data.";
}

<a href=#top>Return to top</a>
<hr><p><img src="smclearhound.gif">&nbsp;<a href="main.html">Back to main NewsHound page</a>
</p>
</body>
</html>