

Web Words Search - Adserver Side Implementation

1) WebWordsSearch.java - Servlet that redirects the user to a search engine that enables the user to search the web for a word from within Eudora

Overview :

i) User searches for a word within Eudora

ii) Eudora contacts the Server using variations of the following URL

<http://jump.eudora.com/jump.cgi?action=search&query=test&platform=WIN&product=Eudora&la=English&distributorID=99999&mode=PAY&version=6.1>

All other parameters passed like 'platform' and 'product' etc except for the 'query' are not required.

iii) Jumptable redirects any incoming request from the client with 'action=search' to the WebWordsSearch Servlet.

iv) Servlet redirects the query to a Search Engine.

v) Servlet logs the search word along with some information related to Eudora in a table.

Implementation :

i) Multi Thread Model

This enables the Servlet to serve multiple customers at the same time. The Servlet makes sure that in spite of several customers making requests at the same time, reads/writes to common data structures and tables are synchronized as required.

ii) Cache

Since DB operations are very expensive, the Servlet does not write to the database each time the user searches for a word. Instead we cache the records in memory and write it out periodically to the database. To decide when to write to the database we check if the size of the data structure in memory goes beyond a thresh hold value (currently it is set to 50) or if the records have been in memory beyond a certain interval of time (dump interval is currently set to 60 minutes). Depending on which condition gets satisfied first of the two, the records from memory are written out to the database. The Servlet also dumps all records in memory to the table when Apache is restarted and the Servlet is destroyed. Thus no records in memory that do not satisfy the above two conditions are lost when the Servlet is destroyed.

iii) Data Structure

ArrayList is used. ArrayLists are preferred over Vectors, since Vectors are synchronized data structures and all of its operation are synchronized even if not required. This slows down the performance when there are multiple users accessing the Vector at the same time. ArrayLists in spite of behaving exactly like Vectors are not synchronized data structures. So the Servlet does the synchronization as and when required. The main operations on the ArrayList are insertions when the user makes a request and then deletions when the records are written out from the ArrayList to the table. Since we can always have multiple users adding to the end of the ArrayList, the Servlet synchronizes the Insertions into the ArrayList. When the size of the ArrayList is checked to decide if we have to dump the data into a DB, it is not synchronized while the process of actually dumping the data to the DB that causes the records to be deleted from the ArrayList is synchronized. Thus all operations to the ArrayList are not synchronized improving the performance when multiple users access the ArrayList at the same time.

iv) Search Engine

I have used Google as the default search engine for now. In future if we need to rotate between different search engines, we will have to automate the process of rotating between different search engines on a periodic basis.

v) Database connections

The Servlet checks each time if the connection to the database is alive because we cannot be sure that we will get requests from clients on a periodic basis and hence the connection might time out. So each time the connection to the database is checked to see if it is alive and if not a new connection is established. Establishing the connection to the database should also be synchronized so that we make sure that at a time only one user get the connection to the database and this prevents multiple connections to the database being established.

2) adserver.websearch.sql - SQL file that creates the websearch table in the adserver database. This websearch table is used to log the user's searches. We only log the search word and NO information that identifies the user is logged.

The websearch table consists of the following fields

Field Extra	Type		Null	Key	Default	
id	int(9)		PRI	NULL	auto_increment	
date	timestamp(14)		YES		NULL	
query	varchar(255)		YES		NULL	
platform	varchar(255)		YES		NULL	
product	varchar(100)		YES		NULL	
la	varchar(10)		YES		NULL	
version	varchar(25)		YES		NULL	
distributorID	varchar(255)		YES		NULL	
mode	enum('PAY','FREE','AD','ALL')	YES		NULL		

