50269-015 (3401-003) *Patent*

UNITED STATES PATENT APPLICATION

FOR

ALIAS-FREE CONTENT-INDEXED OBJECT CACHE

INVENTORS:

PETER MATTIS
JOHN PLEVYAK
MATT HAINES
ADAM BEGUELIN
BRIAN TOTTY
DAVID GOURLEY

PREPARED BY:
MCDERMOTT, WILL & EMERY
SUITE 300
99 CANAL CENTER PLAZA
ALEXANDRIA, VA 22314
(703) 518-5100

ABSTRACT OF THE DISCLOSURE

A method for caching information objects is provided. Information objects are stored in portions of a non-volatile storage device called arenas, which are contiguous regions from which space is allocated in parallel. Objects are contiguously allocated within an arena and are mapped to directory tables that provide an efficient search mechanism. Each object is identified by a name key and a content key. The name key is constructed by applying a hash function to the composition of the name or URL of the object along with implicit or explicit context about the request. The content key is constructed by applying a hash function to the entire contents of the object data. Buckets and blocks in the directory tables store tags and subkeys derived from the keys. Since duplicate objects that have different names will hash to the same content key, the cache can detect duplicate objects even though they have different names, and store only one copy of the object. As a result, cache storage usage is dramatically reduced, and tracking object aliases is not required. The disclosure also encompasses a computer apparatus, computer program product, and computer data signal embodied in a carrier wave that are configured similarly.

FIELD OF THE INVENTION

The present invention relates to information delivery, and relates more specifically to a cache for information objects that are to be delivered efficiently and at high speed over a network to a client.

BACKGROUND OF THE INVENTION

Several important computer technologies rely, to a great extent, upon rapid delivery of information from a central storage location to remote devices. For example, in the client/server model of computing, one or more servers are used to store information. Client computers or processes are separated from the servers and are connected to the servers using a network. The clients request information from one of the servers by providing a network address of the information. The server locates the information based on the provided network address and transmits it over the network to the client, completing the transaction.

The World Wide Web is a popular application of the client/server computing model. Figure 1 is a simplified block diagram of the relationship between elements used in a Web system. One or more web clients 10a, 10b, each of which is a computer or a software process such as a browser program, are connected to a global information network 20 called the Internet, either directly or through an intermediary such as an Internet Service Provider, or an online information service.

A web server 40 is likewise connected to the Internet 20 by a network link 42. The web server 40 has one or more internet network addresses and textual host names, associated in an agreed-upon format that is indexed at a central Domain

Name Server (DNS). The server contains multimedia information resources, such as documents and images, to be provided to clients upon demand. The server 40 may additionally or alternatively contain software for dynamically generating such resources in response to requests.

The clients 10a, 10b and server 40 communicate using one or more agreed-upon protocols that specify the format of the information that is communicated. A client 10a looks up network address of a particular server using DNS and establishes a connection to the server using a communication protocol called the Hypertext Transfer Protocol (HTTP). A Uniform Resource Locator (URL) uniquely identifies each information object stored on or dynamically generated by the server 40. A URL is a form of network address that identifies the location of information stored in a network.

A key factor that limits the performance of the World Wide Web is the speed with which the server 40 can supply information to a client via the Internet 20. Performance is limited by the speed, reliability, and congestion level of the network route through the Internet, by geographical distance delays, and by server load level. Accordingly, client transaction time can be reduced by storing replicas of popular information objects in repositories geographically dispersed from the server. Each local repository for object replicas is generally referred to as a cache. A client may be able to access replicas from a topologically proximate cache faster than possible from the original web server, while at the same time reducing Internet server traffic.

In one arrangement, as shown in Figure 1, the cache is located in a proxy server 30 that is logically interposed between the clients 10a, 10b and the server 40.

The proxy server provides a "middleman" gateway service, acting as a server to the client, and a client to the server. A proxy server equipped with a cache is called a caching proxy server, or commonly, a "proxy cache".

The proxy cache 30 intercepts requests for resources that are directed from the clients 10a, 10b to the server 40. When the cache in the proxy 30 has a replica of the requested resource that meets certain freshness constraints, the proxy responds to the clients 10a, 10b and serves the resource directly. In this arrangement, the number and volume of data transfers along the link 42 are greatly reduced. As a result, network resources or objects are provided more rapidly to the clients 10a, 10b.

A key problem in such caching is the efficient storage, location, and retrieval of objects in the cache. This document concerns technology related to the storage, location, and retrieval of multimedia objects within a cache. The object storage facility within a cache is called a "cache object store" or "object store".

To effectively handle heavy traffic environments, such as the World Wide Web, a cache object store needs to be able to handle tens or hundreds of millions of different objects, while storing, deleting, and fetching the objects simultaneously. Accordingly, cache performance must not degrade significantly with object count. Performance is the driving goal of cache object stores.

Finding an object in the cache is the most common operation and therefore the cache must be extremely fast in carrying out searches. The key factor that limits cache performance is lookup time. It is desirable to have a cache that can determine whether an object is in the cache (a "hit") or not (a "miss") as fast as possible. In past approaches, caches capable of storing millions of objects have been stored in

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.