US007158961B1

(12) **United States Patent**
Charikar

(10) **Patent No.: US 7,158,961 B1**
(45) **Date of Patent: Jan. 2, 2007**

(54) **METHODS AND APPARATUS FOR ESTIMATING SIMILARITY**

(75) Inventor: **Moses Samson Charikar**, Princeton, NJ (US)

(73) Assignee: **Google, Inc.**, Mountain View, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 611 days.

(21) Appl. No.: **10/029,883**

(22) Filed: **Dec. 31, 2001**

(51) **Int. Cl.**
G06F 7/04 (2006.01)

(52) **U.S. Cl.** **707/2**; 707/102

(58) **Field of Classification Search** 707/1–3, 707/5, 10, 100, 102, 104.1; 708/200, 422–424; 345/418, 419, 427; 382/10, 154, 181, 190, 382/191, 209, 217–220, 276–278
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,067,152	A *	11/1991	Kisor et al.	348/422.1
5,101,475	A *	3/1992	Kaufman et al.	345/424
5,469,354	A *	11/1995	Hatakeyama et al.	707/3
5,612,865	A *	3/1997	Dasgupta	700/79
5,794,178	A *	8/1998	Caid et al.	704/9
5,806,061	A *	9/1998	Chaudhuri et al.	707/3
6,061,734	A *	5/2000	London	709/238
6,134,532	A *	10/2000	Lazarus et al.	705/14
6,349,296	B1 *	2/2002	Broder et al.	707/3
6,603,470	B1 *	8/2003	Deering	345/419

OTHER PUBLICATIONS

Moses Sampspn Charikar, "Algorithms for Clustering Problems", 2001, Stanford University, vol. 62/01-B of Dissertation Abstracts International.*

SRC Technical Note; 1997-015; Jul. 25, 1997; "Syntactic Clustering of the Web"; Andrei Z. Broder et al.; pp. 1-14; Digital Equipment Corporation http://gatekeeper.dec.com/pub/DEC/SRC/technical-notes/SRC-1997-015-html/.

"Similarity Search in High Dimensions via Hashing"; Aristides Gionis et al.; Department of Computer Science; Stanford University; pp. 518-529; 1999.

"Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality (preliminary version)"; Piotr Indyk et al.; Department of Computer Science; Stanford University; Jul. 21, 1999; pp. 1-13 and i-vii.

"Chapter 26—Improved approximation algorithms for network design problems"; M.X. Goemans et al.; pp. 223-232.

"Approximation Algorithms for Classification Problems with Pairwise Relationships: Metric Labeling and Markov Random Fields"; Jon Kleinberg et al.

"Scalable Techniques for Clustering the Web"; Taher H. Haveliwala et al.

"Efficient Search for Approximate Nearest Neighbor in High Dimensional Spaces"; Eyal Kushilevitz et al.; pp. 1-17.

"On the resemblance and containment of documents"; Andrei Z. Broder; Digital Systems Research Center; Palo Alto, CA; pp. 1-9.

* cited by examiner

Primary Examiner—Greta Robinson
Assistant Examiner—Harold E. Dodd, Jr.
(74) *Attorney, Agent, or Firm*—Harrity Snyder LLP

(57) **ABSTRACT**

A similarity engine generates compact representations of objects called sketches. Sketches of different objects can be compared to determine the similarity between the two objects. The sketch for an object may be generated by creating a vector corresponding to the object, where each coordinate of the vector is associated with a corresponding weight. The weight associated with each coordinate in the vector is multiplied by a predetermined hashing vector to generate a product vector, and the product vectors are summed. The similarity engine may then generate a compact representation of the object based on the summed product vector.

24 Claims, 4 Drawing Sheets

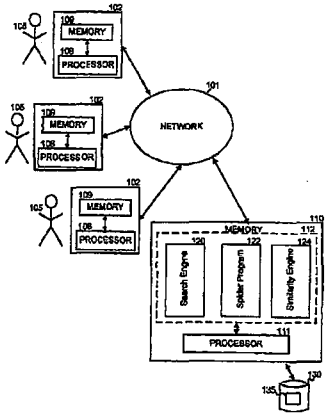

EXHIBIT 2106
Facebook, Inc. et al.
v.

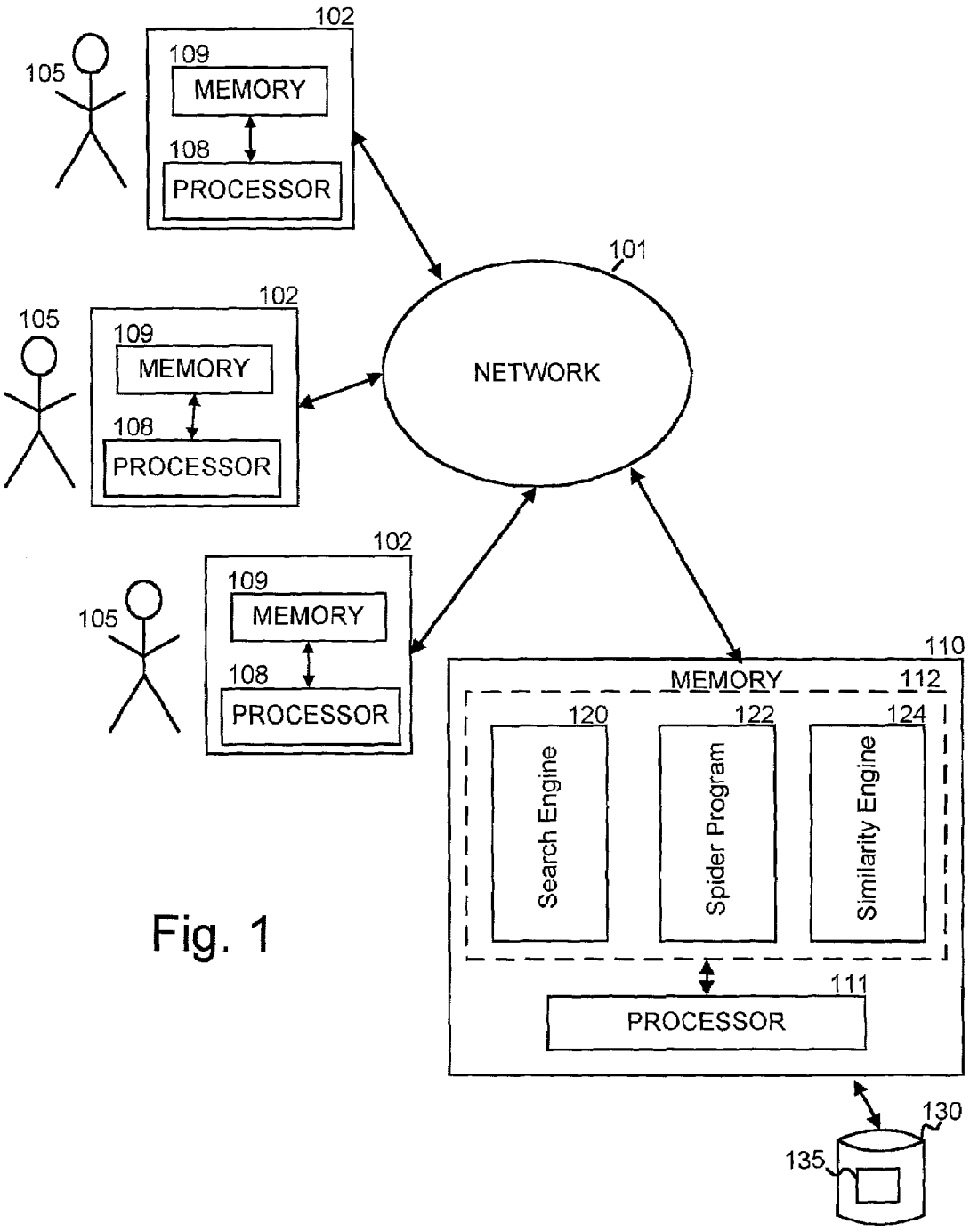

Fig. 1

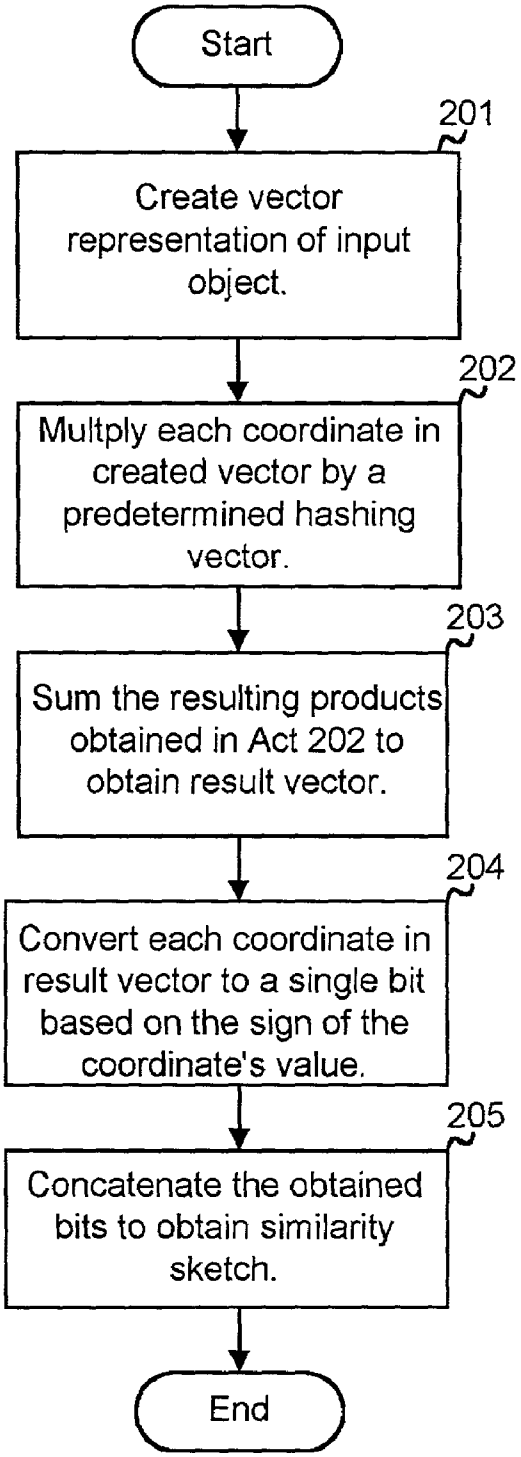

Fig. 2

coordinate	value
four	1
score	1
and	1
seven	1
years	1
ago	1

301

Fig. 3A

coordinate	value
four	0.3
score	0.4
and	0.05
seven	0.3
years	0.4
ago	0.2

302

Fig. 3B

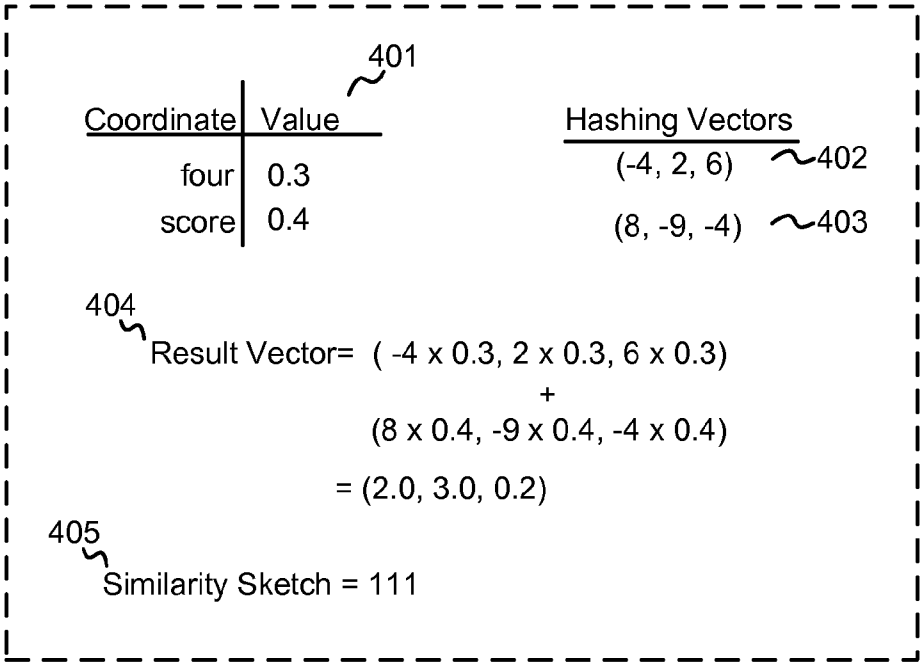

Fig. 4

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.