US006985950B1

(12) **United States Patent**
Hanson et al.

(10) **Patent No.: US 6,985,950 B1**
(45) **Date of Patent: Jan. 10, 2006**

(54) **SYSTEM FOR CREATING A SPACE-EFFICIENT DOCUMENT CATEGORIZER FOR TRAINING AND TESTING OF AUTOMATIC CATEGORIZATION ENGINES**

(75) Inventors: **Andrew Hanson**, Portland, OR (US); **David B. Lamkins**, Portland, OR (US)

(73) Assignee: **Microsoft Corporation**, Redmond, WA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 919 days.

(21) Appl. No.: **09/800,766**

(22) Filed: **Mar. 6, 2001**

(51) **Int. Cl.**
G06F 17/30 (2006.01)

(52) **U.S. Cl.** **709/226**; 709/228; 707/3; 707/4

(58) **Field of Classification Search** 709/219, 709/224, 226, 227, 228, 245; 707/10, 3, 707/105, 4, 5; 713/171
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,461,698 A 10/1995 Schwanke et al.
5,640,468 A 6/1997 Hsu
5,652,829 A 7/1997 Hong
5,657,424 A 8/1997 Farrell et al.
5,706,507 A 1/1998 Schloss
5,708,822 A 1/1998 Wical
5,734,796 A 3/1998 Pao
5,768,580 A 6/1998 Wical
5,809,499 A 9/1998 Wong et al.
5,812,995 A 9/1998 Sasaki et al.
5,835,905 A 11/1998 Pirolli et al.
5,867,799 A 2/1999 Lang et al.
5,870,744 A 2/1999 Sprague
5,911,043 A 6/1999 Duffy et al.
5,920,864 A 7/1999 Zhao
6,055,540 A 4/2000 Snow et al.
6,058,205 A 5/2000 Bahl et al.
6,073,137 A * 6/2000 Brown et al.
6,101,515 A 8/2000 Wical et al.
6,101,527 A * 8/2000 Lejeune et al. 709/201
6,128,613 A 10/2000 Wong et al.
6,161,130 A 12/2000 Horvitz et al.
6,163,778 A 12/2000 Fogg et al.
6,233,575 B1 5/2001 Agrawal et al.
6,249,785 B1 6/2001 Paepke
6,252,988 B1 6/2001 Ho
6,285,999 B1 9/2001 Page
6,370,543 B2 * 4/2002 Hoffert et al. 707/104.1
6,430,558 B1 8/2002 Delano
6,473,753 B1 10/2002 Katariya et al.
6,507,843 B1 1/2003 Dong et al.
6,519,580 B1 2/2003 Johnson et al.
6,526,442 B1 * 2/2003 Stupek et al. 709/224
6,592,627 B1 * 7/2003 Agrawal et al.
6,604,114 B1 8/2003 Toong et al.
6,658,463 B1 * 12/2003 Dillon et al. 709/219

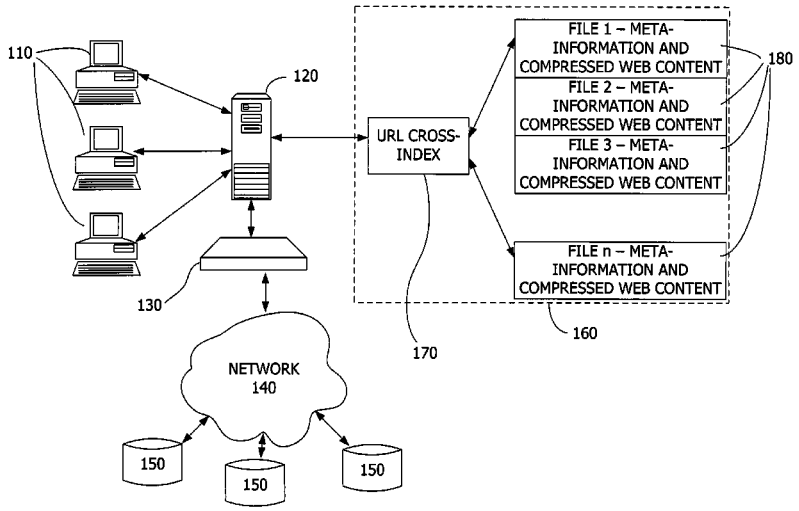
(Continued)

FOREIGN PATENT DOCUMENTS

WO WO 01/33413 A1 5/2001

Primary Examiner—Zarni Maung
Assistant Examiner—Khanh Quang Dinh
(74) *Attorney, Agent, or Firm*—Senniger Powers

(57) **ABSTRACT**

A method of receiving a network resource address and creating a copy of the network resource and storing the created copy in a database is provided. This creation and storage of a copy of the network resource is performed when there does not exist, in the database, information derived from the resource pointed to by the network address.

29 Claims, 8 Drawing Sheets