(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 September 2004 (16.09.2004)

PCT

(10) International Publication Number
WO 2004/079514 A2

(51) International Patent Classification[7]: G06F

(21) International Application Number:
PCT/US2004/005501

(22) International Filing Date: 25 February 2004 (25.02.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/378,463 3 March 2003 (03.03.2003) US

(71) Applicants *(for all designated States except US)*: **MICROSOFT CORPORATION** [US/US]; One Microsoft Way, Redmond, WA 98052 (US). **ROUNTHWAITE, Robert, L.** [US/US]; 4148 287th Avenue SE, Fall City, WA 98024 (US). **HECKERMAN, David, E.** [US/US]; 648 W. Lake Sammamish Lane NE, Bellevue, WA 98008 (US). **MEHR, John, D.** [US/US]; 3624 Whitman Avenue N. #3, Seattle, 98103 (US). **HOWELL, Nathan, D.** [US/US]; 939 N. 105th Street, Apt. A, Seattle, WA 98133 (US). **RUPERSBURG, Micah, C.** [US/US]; 417 Pine Street, #209, Seattle, WA 98122 (US). **SLAWSON, Dean, A.** [US/US]; 3343 W. Ames Lake Drive NE, Redmond, WA 98053 (US). **GOODMAN, Joshua, T.** [US/US]; 17424 NE 38th Street, Redmond, 98052 (US).

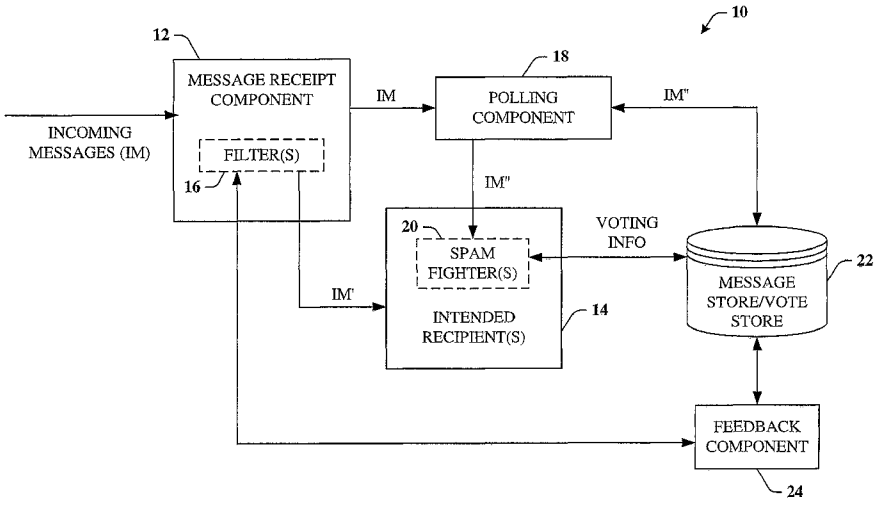
(74) Agents: **AMIN, Himanshu, S.** et al.; Amin & Turocy, LLP, 1900 E. 9th Street, 24th Floor, National City Center, Cleveland, OH 44114 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR,

[Continued on next page]

(54) Title: FEEDBACK LOOP FOR SPAM PREVENTION



(57) Abstract: The subject invention provides for a feedback loop system and method that facilitate classifying items in connection with spam prevention in server and/or client-based architectures. The invention makes uses of a machine-learning approach as applied to spam filters, and in particular, randomly samples incoming email messages so that examples of both legitimate and junk/spam mail are obtained to generate sets of training data. Users which are identified as spam-fighters are asked to vote on whether a selection of their incoming email messages is individually either legitimate mail or junk mail. A database stores the properties for each mail and voting transaction such as user information, message properties and content summary, and polling results for each message to generate training data for machine learning systems. The machine learning systems facilitate creating improved spam filter(s) that are trained to recognize both legitimate mail and spam mail and to distinguish between them.

WO 2004/079514 A2

GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.