

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
2 December 2004 (02.12.2004)

PCT

(10) International Publication Number  
**WO 2004/105332 A3**

(51) International Patent Classification<sup>7</sup>: **H04L 12/58**,  
G06F 17/60

(21) International Application Number:  
PCT/US2004/015383

(22) International Filing Date: 14 May 2004 (14.05.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/471,242 15 May 2003 (15.05.2003) US

(71) Applicant (for all designated States except US):  
**BRIGHTMAIL, INC.** [US/US]; 301 Howard Street,  
18th Floor, San Francisco, CA 94105 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **GLEESON, Matt**;  
\*\* (US). **HOOGSTRATE, David**; \*\* (US). **JENSEN,**  
**Sandy**; \*\* (US). **MANTEL, Eli**; \*\* (US). **MEDLAR,**  
**Art**; \*\* (US). **SCHNEIDER, Ken**; \*\* (US).

(74) Agents: **MALLIE, Michael, J.** et al.; Blakely, Sokoloff,  
Taylor & Zafman LLP, 7th floor, 12400 Wilshire Boule-  
vard, Los Angeles, CA 90025 (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, NA, 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, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI,  
SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,  
GW, ML, MR, NE, SN, TD, TG).

**Published:**

- with international search report
- before the expiration of the time limit for amending the  
claims and to be republished in the event of receipt of  
amendments

(88) Date of publication of the international search report:  
10 March 2005

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

(54) Title: METHOD AND APPARATUS FOR FILTERING EMAIL SPAM BASED ON SIMILARITY MEASURES

(57) Abstract: A method and system for filtering email spam based on similarity measures are described. In one embodiment, the method includes receiving an incoming email message, generating data characterizing the incoming email message based on the content of the incoming email message, and comparing the generated data with a set of data characterizing spam messages. The method further includes determining whether a resemblance between the data characterizing the incoming email message and any data item from the set of data characterizing spam messages exceeds a threshold.



WO 2004/105332 A3

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US2004/015383

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 H04L12/58 G06F17/60

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 H04L G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, INSPEC, COMPENDEX

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6 199 103 B1 (HORI MASAKI ET AL) 6 March 2001 (2001-03-06)	1-3,6,7, 9,12-14, 16, 19-21, 24,25, 27,53-79
Y	abstract column 6, line 35 - column 7, line 5 column 7, line 10 - line 42 ----- -/--	4,5,8, 10,11, 15,17, 18,22, 23,26, 28-52

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

22 November 2004

Date of mailing of the international search report

28.01.2005

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Poppe, F

## INTERNATIONAL SEARCH REPORT

Inter. Application No  
PCT/US2004/015383

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 01/46872 A (PACE MARK RAYMOND ; TALLEY BROOKS CASH (US)) 28 June 2001 (2001-06-28)	1-3,6,7, 9,12-14, 16, 19-21, 24,25, 27,53-79
Y	page 3, line 35 - page 5, line 15 page 7, line 19 - page 8, line 3 -----	4,5,8, 10,11, 15,17, 18,22, 23,26, 28-52
P,X	US 2003/195937 A1 (BHATIA SURENDRA ET AL) 16 October 2003 (2003-10-16)  abstract -----	1,9,13, 16,19, 21,24, 27,53, 65,74,77
A	DAMASHEK M: "GAUGING SIMILARITY WITH N-GRAMS: LANGUAGE-INDEPENDENT CATEGORIZATION OF TEXT" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, vol. 267, no. 5199, 10 February 1995 (1995-02-10), pages 843-848, XP000579827 ISSN: 0036-8075 abstract page 843, right-hand column, line 24 - page 844, left-hand column, line 12 -----	6,7,12, 53-79
Y	JOHN GRAHAM-CUMMING: "The Spammer's Compendium" 'Online! 17 January 2003 (2003-01-17), XP002305731 SPAM CONFERENCE 2003, CAMBRIDGE Retrieved from the Internet: URL: <a href="http://popfile.sourceforge.net/SpamConference011703.pdf">http://popfile.sourceforge.net/SpamConference011703.pdf</a> 'retrieved on 2004-11-15! page 6 - page 12 page 14 - page 15 -----	4,5,8, 10,11, 15,17, 18,22, 23,26, 28-52
A	GRAHAM P: "THE FUTURE OF SPAM" COMPUTER SECURITY JOURNAL, CSI COMPUTER SECURITY INSTITUTE, vol. 19, no. 1, January 2003 (2003-01), pages 1-5, XP008034480 ISSN: 0277-0865 page 3, right-hand column, line 42 - page 4, left-hand column, line 21 -----	1-79

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US2004/015383

### Box II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

### Box III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

#### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-3,6,7,9,12-14,16,19-21,24,25,27,53-79

Methods, system, apparatus and computer-readable medium for determining whether an email message is spam based on the resemblance between its content and the content of email messages that are known to be spam.

---

2. claims: 4-5 as dependent on claim 1, 8 as dependent on claim 3, 10-11 as dependent on claim 9, 15 as dependent on claim 13, 17-18 as dependent on claim 16, 22-23 as dependent on claim 21, 26 as dependent on claim 24, 28-29 as dependent on claim 27, 30-52

Method, system and computer-readable medium for removing noise that was added to a spam message in order to hide its spam nature from a spam filter that bases its decisions on content resemblance.

---

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US2004/015383

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
US 6199103	B1	06-03-2001	JP	11015756 A	22-01-1999
WO 0146872	A	28-06-2001	US	6460050 B1	01-10-2002
			AU	4525901 A	03-07-2001
			EP	1242921 A1	25-09-2002
			JP	2004500761 T	08-01-2004
			WO	0146872 A1	28-06-2001
US 2003195937	A1	16-10-2003	NONE		