



- (51) **International Patent Classification:**
G06F 17/00 (2006.01)
- (21) **International Application Number:**
PCT/US2013/042223
- (22) **International Filing Date:**
22 May 2013 (22.05.2013)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/651,703 25 May 2012 (25.05.2012) US
13/804,096 14 March 2013 (14.03.2013) US
- (71) **Applicant:** CRIMSON HEXAGON, INC. [US/US]; 155 Seaport Blvd., Boston, MA 02210 (US).
- (72) **Inventors:** FIRAT, Aykut; 2 Earhart Street, Unit 804, Cambridge, Massachusetts 02141 (US). BROOKS, Mitchell; 483 Beacon Street, Apt. 63, Boston, Massachusetts 02115 (US). BINGHAM, Christopher; 65 Foster Street, Cambridge, Massachusetts 02138 (US). HERDAG-DELEN, Amac; 201 Flynn Avenue #6, Mountain View, California 94043 (US). KING, Gary; 9 Beecher Road, Brookline, Massachusetts 02445 (US).
- (74) **Agents:** CAHILL, Ronald, E. et al.; Nutter Mcclennen & Fish LLP, Seaport West, 155 Seaport Boulevard, Boston, MA 02210-2604 (US).

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

(84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(88) **Date of publication of the international search report:**

16 January 2014

(54) **Title:** SYSTEMS AND METHODS FOR CALCULATING CATEGORY PROPORTIONS

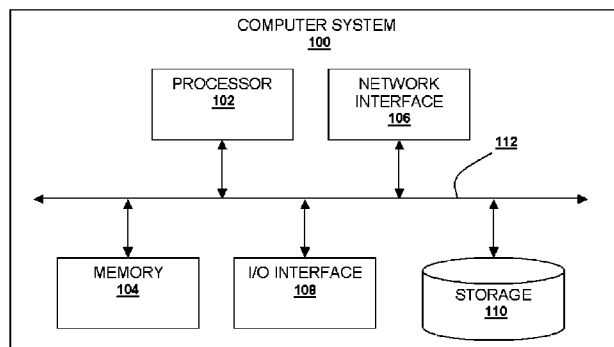


FIGURE 1

(57) **Abstract:** Systems and methods are provided for classifying text based on language. A computer-implemented method includes receiving a training set of elements, each element in the training set being assigned to one of a plurality of categories and having one of a plurality of content profiles associated therewith; receiving a population set of elements, each element in the population set having one of the plurality of content profiles associated therewith; and calculating using at least one of a stacked regression algorithm, a bias formula algorithm, a noise elimination algorithm, and an ensemble method consisting of a plurality of algorithmic methods the results of which are averaged, based on the content profiles associated with and the categories assigned to elements in the training set and the content profiles associated with the elements of the population set, a distribution of elements of the population set over the categories.

