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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 February 2011 (10.02.2011)

PCT

(10) International Publication Number
WO 2011/017155 A4

(51) International Patent Classification:
G06F 17/30 (2006.01)

(21) International Application Number:
PCT/US2010/043516

(22) International Filing Date:
28 July 2010 (28.07.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/229,216	28 July 2009 (28.07.2009)	US
61/236,490	24 August 2009 (24.08.2009)	US
12/844,810	27 July 2010 (27.07.2010)	US

(71) Applicant *(for all designated States except US)*: **FTI TECHNOLOGY LLC** [US/US]; 500 East Pratt Street, Suite 1400, Baltimore, Maryland 21202 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KNIGHT, William, C.** [US/US]; 1851 Edna Place, Bainbridge Island, Washington 98110 (US). **NUSSBAUM, Nicholas, I.** [US/US]; 8114 41st Ave. SW, Seattle, Washington 98136 (US). **CONWELL, John, W.** [US/US]; 16052 48th Ave South, Tukwila, Washington 98188 (US).

(74) Agents: **WITTMAN, Krista A.** et al.; 500 Union Street, Suite 1005, Seattle, Washington 98101 (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, 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, PE, PG, PH, PL, PT, RO, RS, RU, 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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

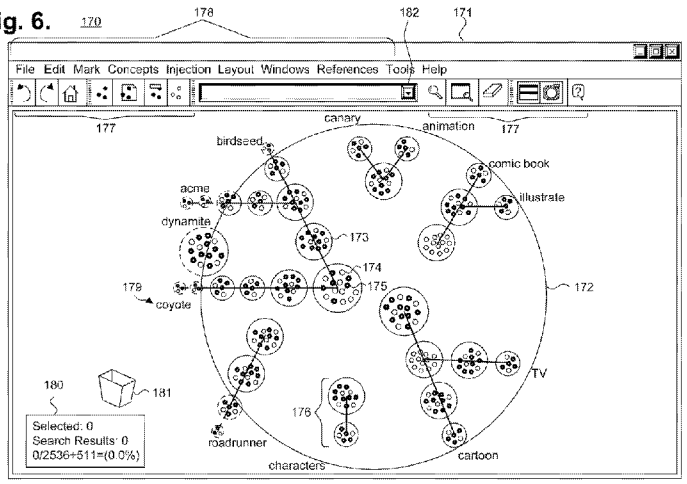
— *with international search report (Art. 21(3))*

— *with amended claims (Art. 19(1))*

Date of publication of the amended claims: 31 March 2011

(54) **Title**: DISPLAYING RELATIONSHIPS BETWEEN CONCEPTS TO PROVIDE CLASSIFICATION SUGGESTIONS VIA INCLUSION

Fig. 6.



(57) **Abstract**: A system (10) and method (50) for displaying relationships between concepts (14c, 14d) to provide classification suggestions via inclusion is provided. A set of reference concepts (14d) each associated with a classification code is designated. One or more of the reference concepts (14d) are combined with a set of uncoded concepts (14c). Clusters of the uncoded concepts (14c) and the one or more reference concepts (14d) are generated. Relationships between the uncoded concepts (14c) and the one or more reference concepts (14d) in at least one cluster (173) are visually depicted as suggestions for classifying the uncoded concepts (14c) in that cluster (173).

WO 2011/017155 A4

AMENDED CLAIMS
received by the International Bureau on 01 February 2011 (01.02.2011)

1. A method (50) for displaying relationships between concepts (14c, 14d) to provide classification suggestions via inclusion, comprising the steps of:

designating a set of reference concepts (14d) each associated with a classification code, wherein each concept (14d) comprises nouns and noun phrases with common semantic meaning that are extracted from a set of documents (14a);

combining a subset of the reference concepts (14d) with a set of uncoded concepts (14c);

generating clusters (173) of the uncoded concepts (14c) and the reference concepts (14d) subset; and

visually depicting relationships between the uncoded concepts (14c) and one or more reference concepts (14d) in at least one cluster (173) as suggestions for classifying the uncoded concepts (14c) in that cluster (173), wherein the steps are performed by a suitably programmed computer.

2. A method (50) according to Claim 1, further comprising:

determining a classification code for at least one of the uncoded concepts (14c) in the at least one cluster (173), comprising:

comparing the at least one uncoded concept (14c) to a neighborhood of reference concepts (14d) in the at least one cluster (173), wherein the neighborhood comprises a predetermined number of the reference concepts (14d) that have a closest distance to the uncoded concept (14c); and

selecting the classification code based on the comparison of the at least one uncoded concept (14c) and reference concept neighborhood.

3. A method (50) according to Claim 1, further comprising:

automatically classifying the at least one uncoded concept (14c) by assigning the classification code to that uncoded concept (14c).

4. A method (50) according to Claim 3, further comprising:

identifying the documents (14a) associated with the classified uncoded concept (14c); and

classifying the associated documents (14a) by assigning a classification code to each of the documents (14a).

5. A method (50) according to Claim 4, wherein the documents (14a) are identified using a matrix comprising a mapping of concepts and related documents (14a).

6. A method (50) according to Claim 2, further comprising:

providing the classification code as a suggestion for classifying the at least one uncoded concept (14c) in the at least one cluster (173).

7. A method (50) according to Claim 6, further comprising:

providing a confidence level for the classification code suggestion.

8. A method (50) according to Claim 7, wherein the classification code is provided when the confidence level exceeds a predetermined threshold.

9. A method (50) according to Claim 1, further comprising:

generating the set of reference concepts (14d), comprising at least one of:

identifying dissimilar uncoded concepts (14c) for a document review project and assigning a classification code to each of the dissimilar uncoded concepts (14c) for inclusion in the reference concepts (14d) set; and

clustering uncoded concepts (14c) for a document review project, selecting one or more of the uncoded concepts (14c) in one or more of the clusters (173), and assigning a classification code to each of the selected uncoded concepts (14c) for inclusion in the reference concepts (14d) set.

10. A method (50) according to Claim 1, wherein the subset of reference concepts (14d) is determined via at least one of predefined, arbitrary, or customized selection.

11. A system (10) for displaying relationships between concepts to provide classification suggestions via inclusion, comprising:

a set of reference concepts (14d) each associated with a classification code, wherein each concept comprises nouns and noun phrases with common semantic meaning that are extracted from a set of documents (14a);

a clustering module (33) to combine a subset of the reference concepts (14d) with a set of uncoded concepts (14c) and to generate clusters (173) of the uncoded concepts (14c) and the reference concepts (14d) subset;

a display (36) to visually depict relationships between the uncoded concepts (14c) and one or more reference concepts (14d) in at least one cluster (173) as suggestions for classifying the uncoded concepts (14c) in that cluster (173); and

a processor to execute the clustering module.

12. A system (10) according to Claim 11, further comprising:

a classification module (34) to determine a classification code for at least one of the uncoded concepts (14c) in the at least one cluster (173), comprising:

a comparison module to compare the at least one uncoded concept (14c) to a neighborhood of reference concepts (14d) in the at least one cluster (173), wherein the neighborhood comprises a predetermined number of the reference concepts (14d) that have a closest distance to the uncoded concept (14c); and

a selection module to select the classification code based on the comparison of the at least one uncoded concept (14c) and reference concept neighborhood.

13. A system (10) according to Claim 11, wherein the classification module (34) automatically classifies the at least one uncoded concept (14c) by assigning the classification code to that uncoded concept (14c).

14. A system (10) according to Claim 13, wherein the classification module (34) identifies the documents (14a) associated with the classified uncoded concept (14c) and classifies the associated documents (14a) by assigning a classification code to each of the documents (14a).

15. A system (10) according to Claim 14, wherein the documents (14a) are identified using a matrix comprising a mapping of concepts and related documents (14a).

16. A system (10) according to Claim 12, wherein the classification module (34) provides the classification code as a suggestion for classifying the at least one uncoded concept (14c) in the at least one cluster (173).

17. A system (10) according to Claim 16, wherein the classification module (34) provides a confidence level with the classification code suggestion.

18. A system (10) according to Claim 17, wherein the classification code is provided when the confidence level exceeds a predetermined threshold.

19. A system (10) according to Claim 11, further comprising:

a reference module to generate the set of reference concepts (14d), comprising at least one of:

a reference similarity module to identify dissimilar uncoded concepts (14c) for a document review project and to assign a classification code to each of the dissimilar uncoded concepts (14c) for inclusion in the reference concepts (14d) set; and

a reference clustering module to cluster uncoded concepts (14c) for a document review project, to select one or more of the concepts in at least one cluster (173), and to assign a classification code to each of the selected concepts for inclusion in the reference concepts (14d) set.

20. A system (10) according to Claim 11, wherein the subset of reference concepts (14d) is determined via at least one of predefined, arbitrary, or customized selection.