(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété Intellectuelle
Bureau international

(43) Date de la publication internationale
16 juin 2005 (16.06.2005)

PCT

(10) Numéro de publication internationale
WO 2005/055086 A1

(51) Classification internationale des brevets[7] : G06F 17/30

(21) Numéro de la demande internationale :
PCT/FR2004/003017

(22) Date de dépôt international :
25 novembre 2004 (25.11.2004)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
0313907 27 novembre 2003 (27.11.2003) FR

(71) Déposant *(pour tous les États désignés sauf US)* : **ADVESTIGO** [FR/FR]; 1 Avenue de la Terrasse, Bât.5, F-91198 Gif sur Yvette (FR).

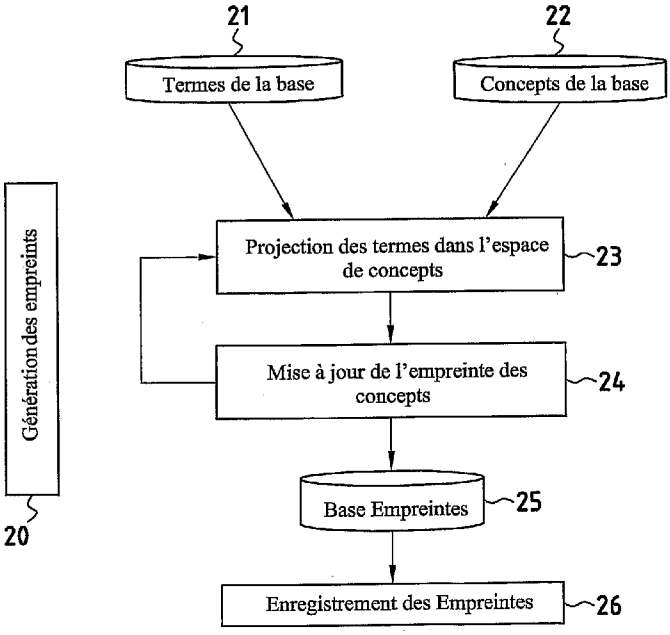
(72) Inventeurs; et

(75) Inventeurs/Déposants *(pour US seulement)* : **ESSAFI, Hassane** [FR/FR]; 97 Rue de Paris, Résidence de l'Yvette - Bât. 1, F-91400 ORSAY (FR). **GUEZOULI, Larbi** [DZ/FR]; Résidence les Fonds Fanettes - Apt. 13, 8 Rue de la Croix Audienne, F-91190 Gif sur Yvette (FR). **SAYAH, Salima** [DZ/FR]; c/o Advestigo, 1 Avenue de la Terrasse - Bât.5, F-91198 Gif sur Yvette (FR). **BEHLOUL, Ali** [DZ/FR]; c/o Advestigo, 1 Avenue de la Terrasse - Bât.5, F-91198 Gif sur Yvette (FR). **MANDRIDAKE, Clarisse**

[Suite sur la page suivante]

(54) **Title:** METHOD FOR INDEXING AND IDENTIFYING MULTIMEDIA DOCUMENTS

(54) **Titre :** PROCEDE D'INDEXATION ET D'IDENTIFICATION DE DOCUMENTS MULTIMEDIAS



20... PRINT GENERATION
21... BASE TERMS
22... BASE CONCEPTS
23... PROJECTION OF TERMS IN CONCEPT SPACE
24... CONCEPT PRINTS UPGRADE
25... PRINT BASE
26... PRINT RECORDING

(57) Abstract: The method for indexing multimedia documents comprises the following steps: (a) identification and extraction, for each document, of terms t_i comprising vectors characterizing the properties of the document; (b) storage of terms t_i in a term base comprising P terms; (c) determination of a maximum number N of desired concepts including the most relevant terms t_i; (d) calculation of the matrix T of distances between the terms t_i of the term base, (e) decomposition of the set P of terms t_i of the term base into N parts P_j (1=j=N) such that P = P1 UP_2 UP_j UP<SB>N</SB>, each part P_j comprising a set of terms t_{ij} and being represented by a concept c_j, terms t_i being divided in such a manner that the furthermost terms are located in distinct parts P_l, P_m and the nearest terms are located in the same part P_l; (f) structuring of the dictionary of concepts and (g) construction of a prints base consisting of the set of concepts c_i representing the terms t_i of said documents, each document being associated with a print which is proper thereto.

(57) Abrégé : Le procédé d'indexation de documents multimédias comprend les étapes suivantes : (a) identification et extraction pour chaque document de termes t_i constitués par des vecteurs caractérisant des propriétés du document ; (b) stockage des termes t_i dans une base de termes comprenant P termes ; (c) détermination d'un nombre maximum N de concepts souhaités regroupant les termes t_i les plus pertinents ; (d) calcul de la matrice T de distances entre les termes t_i de

[Suite sur la page suivante]

WO 2005/055086 A1

[MG/FR]; 18 Rue du Baron de Nivière, F-91140 Villebon sur Yvette (FR). **ESSAFI, Louafi** [MA/FR]; 1 Rue Gabriel Josserand, F-93500 Pantin (FR).

(74) **Mandataires : THEVENET, Jean-Bruno** etc.; Cabinet Beau de Loménie, 158 Rue de l'Université, F-75340 PARIS Cedex 07 (FR).

(81) **États désignés** *(sauf indication contraire, pour tout titre de protection nationale disponible)* **:** 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) **États désignés** *(sauf indication contraire, pour tout titre de protection régionale disponible)* **:** ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, 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).

Publiée :

— *avec rapport de recherche internationale*

— *avant l'expiration du délai prévu pour la modification des revendications, sera republiée si des modifications sont reçues*

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

la base de termes, (e) décomposition de l'ensemble P des termes t_i de la base de termes en N parties P_j ($1 \leq j \leq N$) telles que $P = P1\ UP_2 \ldots UP_j \ldots UP_N$, chaque partie P_j comprenant un ensemble de termes t_{ij} et étant représentée par un concept c_j, les termes t_i étant répartis de telle façon que les termes les plus éloignés se trouvent dans des parties P_l, P_m distinctes et les termes proches se trouvent dans la même partie P_l ; (f) structuration du dictionnaire de concepts et (g) construction d'une base d'empreintes constituée de l'ensemble des concepts c_i représentant les termes t_i des documents, chaque document étant associé à une empreinte qui lui est propre.