A B S T R A C T

A METHOD OF INDEXING AND IDENTIFYING MULTIMEDIA DOCUMENTS

The method of indexing multimedia documents comprises the following steps: a) for each document identifying and extracting terms t_i constituted by vectors characterizing properties of the; b) storing terms t_i in a term base comprising P terms; c) determining a maximum number N of desired concepts that group together the most pertinent terms t_i; d) calculating the matrix T of distances between the terms t_i of the term base; e) decomposing the set P of terms t_i of the term base into N portions P_j ($1 \leq j \leq N$) such that $P = P_1 \cup P_2 \ldots \cup P_j \ldots \cup P_N$, each portion P_j comprising a set of terms t_{ij} and being represented by a concept c_j, the terms t_i being distributed in such a manner that the terms that are farther apart are to be found in distinct portions P_l, P_m, and the terms that are closer together are to be found in the same portion P_l; f) structuring the concept dictionary; and g) constructing a fingerprint base made up of the set of concepts c_i representing the terms t_i of the documents, each document being associated with a fingerprint that is specific thereto.

Translation of the title and the abstract as published by the PCT Authorities, possibly after making changes, ex officio, e.g. under PCT Rules 37.2, 38.2, and/or 48.3.