US 20160026625A1

(19) **United States**

(12) **Patent Application Publication**

Walker

(10) **Pub. No.: US 2016/0026625 A1**

(43) **Pub. Date: Jan. 28, 2016**

(54) **GLOBAL OPTIMIZATION STRATEGIES FOR INDEXING, RANKING AND CLUSTERING MULTIMEDIA DOCUMENTS**

(71) Applicant: **Tapicu Inc**, Long Beach, CA (US)

(72) Inventor: **Reginald L. Walker**, Los Angeles, CA (US)

(73) Assignee: **Tapicu Inc.**, Long Beach, CA (US)

(21) Appl. No.: **14/756,283**

(22) Filed: **Aug. 20, 2015**

Related U.S. Application Data

(63) Continuation-in-part of application No. 13/135,943, filed on Jul. 19, 2011, now Pat. No. 8,825,562.

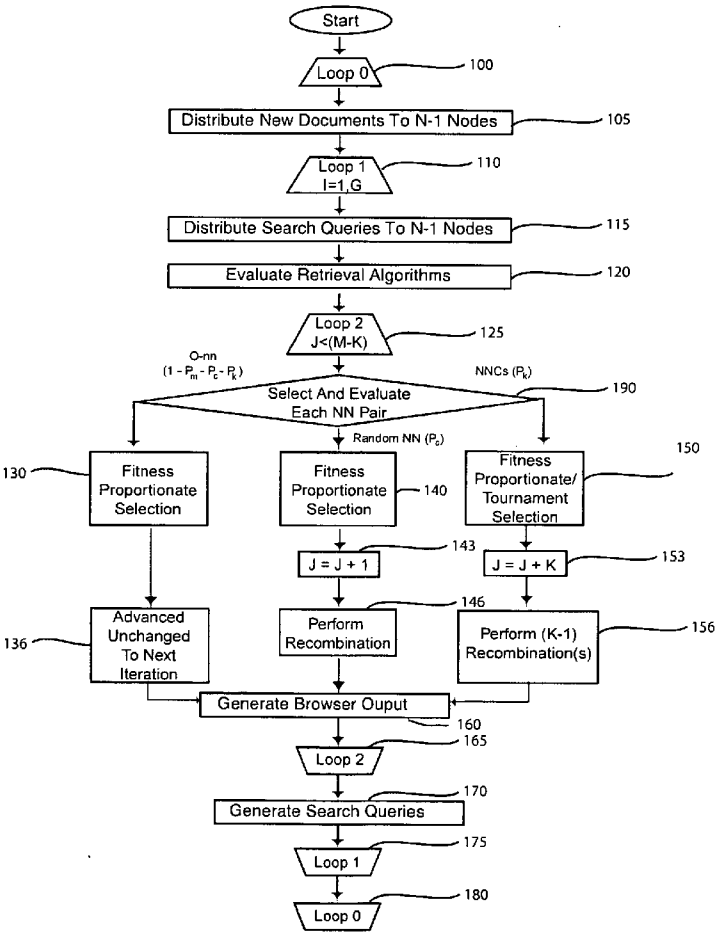
Publication Classification

(51) **Int. Cl.**
G06F 17/30 (2006.01)

(52) **U.S. Cl.**
CPC ***G06F 17/3002*** (2013.01); ***G06F 17/30705*** (2013.01); ***G06F 17/30616*** (2013.01)

(57) **ABSTRACT**

A method for a system that indexes, ranks, and clusters multimedia documents using assessment means, scoring means, stochastic means, and organizing means that optimizes parameter sets comprising of object parameters. The method creates a plurality of individual parameter sets, the parameter sets comprising information sharing system object parameters for describing a model, structures, shape, design, process, search query sets, and dynamic search spaces to be optimized using selective variations, constructive variations, clustering variations, and stochastic variations. The optimizations are guided by document query terms of the search query set object parameter that are initially optimized by assessment means, scoring means, stochastic means, and organizing means that lead to selective variations, constructive variations, clustering variations, and stochastic variations of the parameter sets. The global optimization of parameter sets leads to stochastically improvements to all object parameters by selective variations, constructive variations, clustering variations, and stochastic variations.



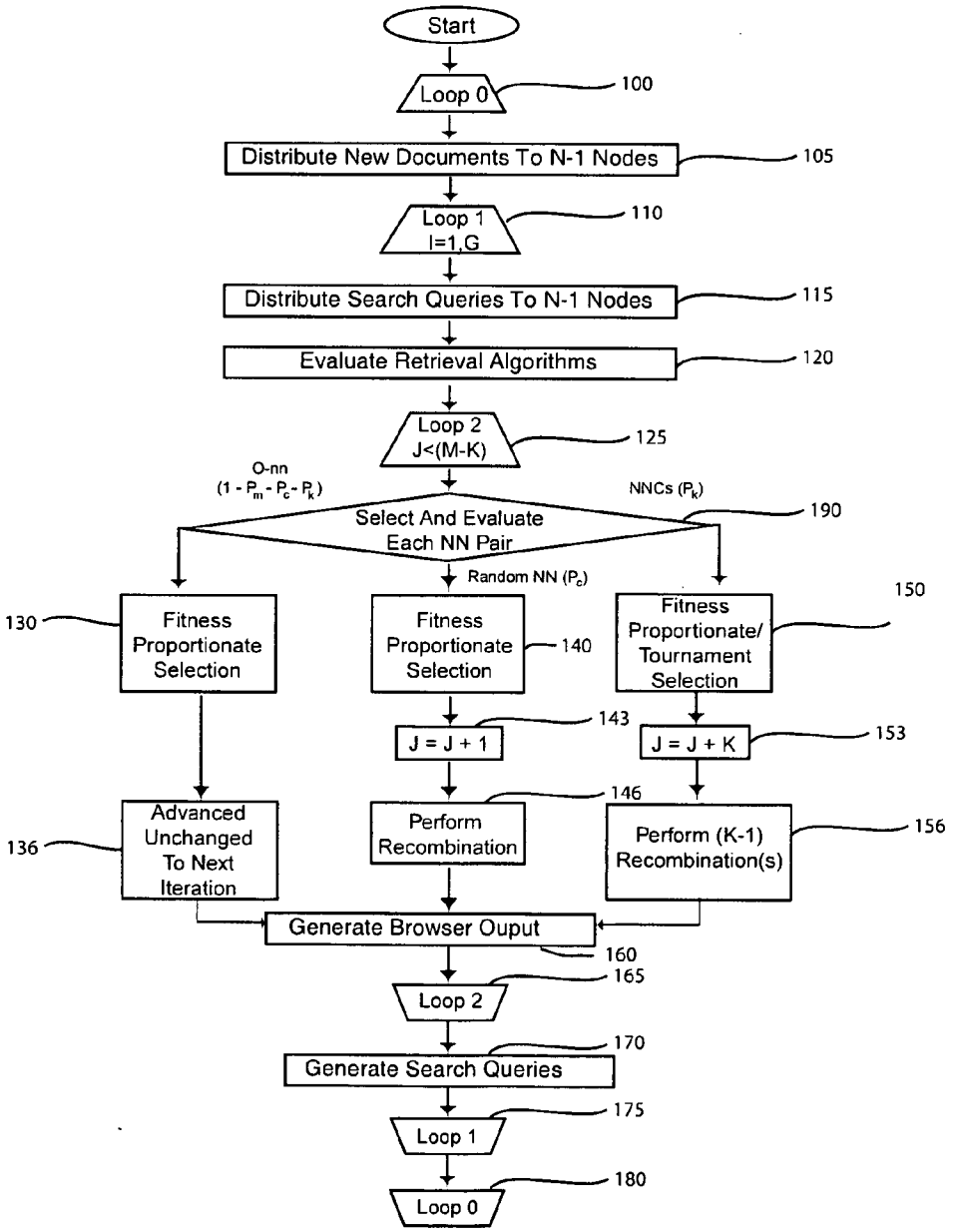

Start
Loop 0
100
Distribute New Documents To N-1 Nodes
105
Loop 1
I=1,G
110
Distribute Search Queries To N-1 Nodes
115
Evaluate Retrieval Algorithms
120
Loop 2
J<(M-K)
125
O-nn
(1 - Pm - Pc - Pk)
NNCs (Pk)
Select And Evaluate Each NN Pair
190
Random NN (Pc)
130
Fitness Proportionate Selection
Fitness Proportionate Selection
140
Fitness Proportionate/ Tournament Selection
150
J = J + 1
143
J = J + K
153
Advanced Unchanged To Next Iteration
136
Perform Recombination
146
Perform (K-1) Recombination(s)
156
Generate Browser Ouput
160
Loop 2
165
Generate Search Queries
170
Loop 1
175
Loop 0
180

FIG. 1

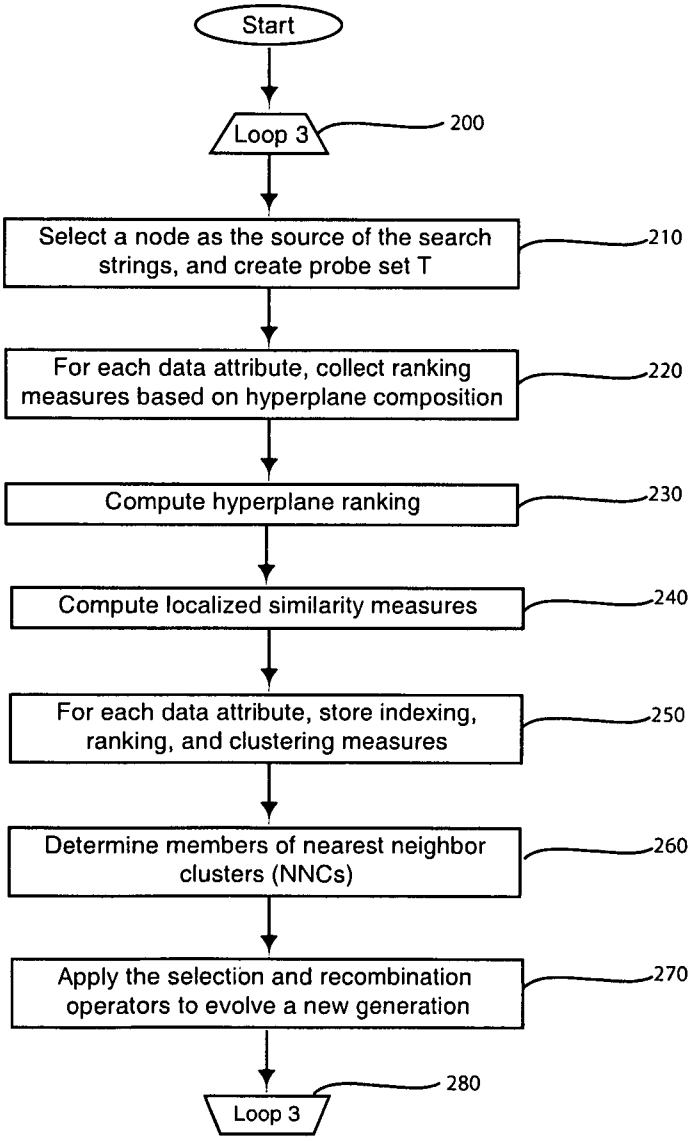

Start
Loop 3
200
Select a node as the source of the search strings, and create probe set T
210
For each data attribute, collect ranking measures based on hyperplane composition
220
Compute hyperplane ranking
230
Compute localized similarity measures
240
For each data attribute, store indexing, ranking, and clustering measures
250
Determine members of nearest neighbor clusters (NNCs)
260
Apply the selection and recombination operators to evolve a new generation
270
Loop 3
280

FIG. 2

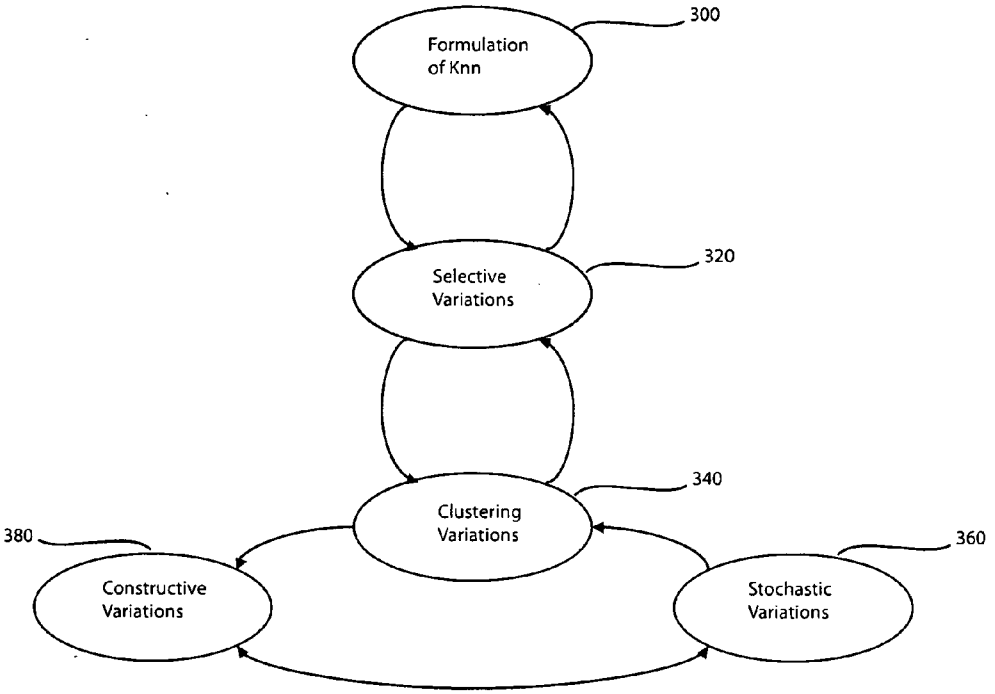

300
Formulation of Knn
320
Selective Variations
340
Clustering Variations
380
Constructive Variations
360
Stochastic Variations

FIG. 3

GLOBAL OPTIMIZATION STRATEGIES FOR INDEXING, RANKING AND CLUSTERING MULTIMEDIA DOCUMENTS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims benefit of provisional patent application No. 62/123,221, filed on Nov. 10, 2014, by the present inventor. This application claims benefit of U.S. patent application Ser. No. 14/545,524, filed on May 15, 2015, by the present inventor. This application claims benefit of continuation-in-part of U.S. patent application Ser. No. 13/135,943, filed on Jul. 19, 2011, now U.S. Pat. No. 8,825,562. Each of the above-referenced applications are incorporated herein by reference in their entirety.

FEDERALLY SPONSORED RESEARCH

[0002] Not Applicable

SEQUENCE LISTING OR PROGRAM

[0003] Not Applicable

BACKGROUND OF THE INVENTION

[0004] The invention relates generally to the optimizing of object parameters for describing a model, structure, shape, design, or process, for an information sharing indexer system. In particular, it relates to the stochastic optimization of evolutionary computation (EC) search strategy parameters for multimedia indexers for information sharing indexer systems such as search engines, data warehouses, and service oriented architectures (SOAs). The field of evolutionary computation encompasses stochastic optimization techniques, such as randomized search strategies, in the form of evolutionary strategies (ES), evolutionary programming (EP), genetic algorithms (GA), classifier systems, evolvable hardware (EHW), and genetic programming (GP).

[0005] There has always been a need to iteratively improve the clustering and ranking of multimedia documents. The stochastic optimization techniques of evolutionary computation (EC) contain mechanisms which enable the representation of certain unique aspects of individual behavior to improve document clustering. Principles of the stochastic optimization techniques of EC can be found for example in Reginald Louis Walker (2003) "*Tocorime Apicu: Design of an Experimental Search Engine Using an Information Sharing Model*", University of California Dissertation, UMI Dissertation Publishing, Ann Arbor, Mich. 48106-1346, which is incorporated by reference herein in its entirety.

[0006] The chief differences among the various types of EC stemming from: 1) the representation of solutions (known as individuals in EC), 2) the design of the variation operators (mutation and/or recombination—also known as crossover), and 3) selection mechanisms. A common strength of these optimization approaches lies in the use of hybrid algorithms derived by combining two or more of the evolutionary search methodologies. The underlying optimization methodologies of EC are used to implement unique stochastic aspects of search strategies that are combined with information retrieval methodologies. This mapping is extended by supplementing the search strategies with finding hidden knowledge in a collection of multimedia documents--related and/or unrelated—using search query sets. Canonical multimedia documents are generated to reduce the workload and storage requirements of the system, resulting in a set of condensed multimedia documents forming the data store. The system continuously repartitions the stored document space among a set of nodes whose goal is to form subclusters of nodes for redistributing the workload. The subclusters are formed by using the information retrieval (IR) algorithm metrics coupled with two or more evolutionary search strategies as the basis of nearest neighbor clusters (NNC) among multimedia indexers. Fitness proportionate and tournament selection in this application forms the basis of nearest neighbor clustering, providing the mechanism for selecting nodes that will share information. Mutations and recombinations are implemented as random change (or multiple changes) of the description of the finite state machine (FSM) according to five different modifications: change of an output symbol, change of a state transition, addition of a state, deletion of a state, or change of the initial state.

OBJECTIVES

[0007] Accordingly, the objectives and advantages of the invention are as follows:

[0008] It is an objective of the present invention to use hybrid algorithms derived by combining one or more of the information retrieval methodologies with one or more of the evolutionary computation search methodologies.

[0009] It is another objective of the present invention is to provide a stochastic selection process that iteratively improves a population of solutions--evolving sets of competing solutions over the space being searched. The components of an optimization application are:

[0010] 1. Terminal set. Input variables or constants.

[0011] 2. Function set. Domain-specific functions that construct potential solutions.

[0012] 3. Fitness measure(s). Function(s) that assign numeric values to the individuals associated with a population (set of solutions that comprise the solution space).

[0013] 4. Algorithm control parameters. Settings dependent on population size and workload redistribution (recombination and mutation) rates.

[0014] 5. Termination criterion. Predicate that uses fitness measures to determine the appropriateness of a population based on tolerances or limits on the number of allowable generations/iterations.

[0015] It is another objective of the present invention to represent solutions as memes to reduce in the computational effort to achieve the periodic optimal document clusters. The fitness of a species (adaptive and iterative grouping of the solutions from selective indexers) can be improved by the non-genetic transmission of cultural information that uses a meme as the transmission mechanism rather than the genetically based gene. The difference between the two includes the fact that genetic transmissions (stochastic selection process) evolve over a period of generations, whereas cultural transmissions result from an educational process.

[0016] It is another objective of the present invention to use a function set that consists of a multimedia parser that works as a two-pass parser. The initial pass occurs as a component of the system that applies document layout analysis for its automated retrieval component. The second pass applies a full set of text-processing modules consisting of syntactic analysis, lexical analysis, layout analysis, and feature recognition. Layout analysis transforms a raw document into an application-specific document by saving the canonical format struc-

tural information as necessary. The syntactic analysis component verifies that the canonical structure adheres to a suitable format. The lexical analysis module is combined with the feature recognition module. These modules remove stop words, identify and record word boundaries, and index words for retrieval. Additionally, this component is responsible for converting hyphenated and sequences of capitalized words into proximity constraints, and case conversions into compressed inverted files.

[0017] It is another objective of the present invention to continuously apply algorithm control parameters to improve the subclustering of documents in distributive applications leading to disjoint nodes for chosen sets of search queries.

[0018] It is another objective of the present invention to continuously adjust the operational parameters required to filter, organize, and index any large-scale data set—information stored on a single computer, a local area network (LAN), and a wide area network (WAN) that encompasses the whole Internet—that may consists of constantly fluctuating information content over relatively short periods of time.

SUMMARY OF THE INVENTION

[0019] The invention is a system and method for indexing, ranking, and clustering multimedia documents using hybrid search strategies and the stochastic optimization techniques of evolutionary computation (EC). These stochastic optimization techniques form the basis of a regulatory mechanism for sharing information document clustering and ranking which leads to the migration of multimedia documents between multimedia indexers. The iterative application of these mechanisms improves the subclustering of multimedia documents in distributive applications leading to disjoint nodes for chosen sets of search queries according to one embodiment.

[0020] In one embodiment, a plurality of individual parameter sets are created wherein the parameter sets comprise information sharing system object parameters for describing a model, structures, shape, design, process, search query sets, and dynamic search spaces to be optimized using selective variations, constructive variations, clustering variations, and stochastic variations. The optimizations are guided by document query terms of the search query set object parameter.

[0021] It is to be understood that both foregoing general description and the following detailed description for the present invention are explempary and explanatory and are extended to provide further explanation of the invention as claimed.

DETAILED DESCRIPTION OF THE DRAWINGS

Figures

[0022] FIG. **1** is a schematic flow diagram of the optimization method of the present invention.

[0023] FIG. **2** is illustrates a flow diagram for optimizing object parameters of parameter sets using assessment means, scoring means, stochastic means, and organizing means to generate ranking scores, indexing scores, and clustering scores

[0024] FIG. **3** is an illustration of the optimization strategies using selective variations, constructive variations, clustering variations, and stochastic variations, in accordance with one or more embodiments.

PREFERRED EMBODIMENTS

[0025] A preferred embodiment of the present invention is now described with reference to the figures where like reference numbers indicate identical or functionally similar elements.

[0026] Some portions of the detailed descriptions that follow are presented in terms of algorithms and symbolic representations of operations on data bits within a computer memory. These algorithmic descriptions and representations are the means used by those skilled in data processing art to most effectively convey the substance of their work to others skilled in the art. Algorithms are here, and generally, conceived to be self- consistence sequence of steps (instructions) leading to desired results. The steps are those requiring physical manipulations of physical quantities.

[0027] Certain aspects of the present invention include process steps and instructions described herein in the form of an algorithm. It should be noted that the process steps and instructions of the present invention could be embodied in software, could be downloaded to reside on and be operated from different platforms used by a variety of operating systems.

[0028] The present invention also relates to an apparatus for performing the operations herein. This apparatus may be specially constructed for the required purposes, or it may comprise a general-purpose computer selectively activated or reconfigured by a computer program stored in a computer. Furthermore, the computers referred to in the specifications may include a single processor or may be architectures employing multiple processors designed for increased computing capability.

[0029] The algorithms and displays presented herein are not inherently related to any particular computer of other apparatus. Various general-purpose systems may also be used with programs in accordance with the teaching herein, or it may prove convenient to construct more specialized apparatus to perform the required method steps. The required structure for a variety of these systems will appear from the description below. In addition, the present invention is not described with reference to any programming language. It will be appreciated that a variety of programming languages may be used to implement the teachings of the present invention as describe herein, and any references below to specific languages are provided for disclosure of enablement and best mode of the present invention.

[0030] In addition, the language used in the specification has been principally selected for readability and instructional purposes, and may not have been selected to delineate or circumscribed the invention subject matter. Accordingly, the disclosure of the present invention is intended to be illustrative, but not limiting, of the scope of the invention, which is set forth in the claims.

[0031] 1. Notational Conventions

a. Applying the EC Search Strategies to Stochastic Information Fluctuations

[0032] FIG. **1** shows an embodiment of the invention which performs the tasks associated with regulating the formulation of NNCs and adapting to information fluctuations. The tasks performed are:

[0033] 1. Periodic partitioning of the multimedia document dataset among indexer nodes **105**

[0034] 2. Generating workload assignments (resulting from fitness proportionate selection steps) for each node **105**

[0035] 3. Distributing dynamic workload assignments **105** and dynamic search query sets **115**

[0036] 4. Formulating NNCs using fitness proportionate selection **190**

[0037] 5. Selecting source of dynamic search query sets **170**

[0038] 6. Repeating step 1 through 5 **100-180**

[0039] The traditional EC approach for the recombination and mutation operators, as well as the normal (steady-state) approach, is restricted to one application per iteration for a single set of solutions. The load-balancing model of the multimedia document indexing system uses the EC recombination operator by restricting information sharing between members of disjoint node sets (species) which are chosen in a process that selects and evaluate each nearest neighbor (NN) pair **190**.

[0040] NNCs **190** can occur as one of three types based on the number of neighborhood seeds: 1) random seeds, 2) multiple seeds, or 3) overlapping seeds. The occurrence of multiple and overlapping seeds enhances the quality of the total cluster's solution space via the modification of the workload assignments of several nodes during one iteration (superstep).

[0041] The iterative formulation of NNCs **125**,**190** was implemented using the notion of an expandable search space which facilitates adaptive subclusters on an iteration-by-iteration basis. The selection process **190** can be applied multiple times **153**, where one node is the NN seed for one or more nodes—thus providing a stochastic hybrid of the recombination and mutation operators **130**,**140**,**150**.

b. Formulation of Nearest Neighbor Clusters (NNCs)

[0042] K-nearest neighbors (K-nn) **190**,**130**,**136** is implanted as the mutation operator when K=0.

[0043] Random NN **190**,**140**,**143**,**146** are implemented as follows: 1) the first node is randomly chosen, and 2) the second node is chosen by incrementing the node ID of the first node **190**, thereby mimicking the ring communication pattern based on the rank in order to determine adjacent nodes. Recombination is applied to the selected nodes **140**,**143**,**146** for each iteration **125-165**. The proportionate fitness method **140** assigns a random number to each neighborhood seed and selects individuals by repeatedly choosing various random numbers until one matches a node's random number.

[0044] Multiple neighborhoods (NNCs) **190**,**150**,**153**,**156** exists when there are at least one or more NNCs in which neighborhoods do not overlap. When a single node is a nearest neighbor of two disjoint NNCs, this node may be selected **150** as a NN one or more times based on the existence of one or more completing nodes in the disjoint neighborhoods. The selection of a node when two or more are present in a single neighborhood occurs via proportionate fitness selection **150**.

[0045] Overlapping neighborhoods **190**,**150**,**153**,**156** occur when two or more NNCs are formed from the seeds overlapping neighborhoods. The selection of one of the NNCs **150** from overlapping of neighborhoods occurs via two "popular" selection methodologies: 1) the proportionate fitness or roulette wheel selection, and 2) the tournament selection. The proportionate fitness method **150** assigns a random number to each node and selects individuals choosing various random numbers which may match an individual's random number.

[0046] The selection processes **190**,**150** for overlapping neighborhoods uses the radius of two or more nodes resulting in possibly K-nn per cluster by performing the following:

[0047] 1. Randomly selects one of the overlapping nodes as the seed of one of the NNCs using the tournament selection method **150**

[0048] 2. Using roulette wheel selection **150**

[0049] a. Randomly selects a node for recombination

[0050] b. Randomly selects a range for recombination

[0051] c. Performs recombination **156** on the two nodes only if they are NN using proportionate fitness method **150**

[0052] 3. If necessary, repeats step 2 **125-165**

The number of iterations **156** a selected node is used for recombination is random—this potentially providing the node with an emulator of the mutation operator **130** (occurring if the selected node was previously selected during an application of the recombination operator). However, the same node may be chosen for two or more iterations with the possibility of swapping previously exchanged recombinations. The system does not advance until k possible recombinations **156** have been completed. The occurrence of overlapping NNCs regulates the recombination rate and the selection rate. The recombination rate and the selection rate use the information retrieval algorithms to generate stochastic metrics for determining nearest neighbor (NN) resulting in the emergence of subclustering within each cluster/subcluster since each meme is maintained throughout this application.

[0053] Another component of the recombination rate and the selection rate stems from overlapping nearest neighbor clusters (NNCs) and is equivalent to sharing information between diverse set of computer processors and/or systems. This phenomenon adds random noise to the whole process by creating, at most K-nn in one component of a superstep based on overlapping NNCs—an event which is beneficial to the prevention of premature convergence and to the incorporation of various optimization techniques such as supersteps and dissassortive mating when selecting nodes from initial subclusters such subspecies A and B. Supersteps resulted from two or more applications of the recombination operator during one iteration (generation) via overlapping NNCs or multiple disjoint NNCs. Dissassortive selection is a results of selecting NN for the recombination operator from a disjoint list of disjoint subcluster members, as in the case of random NN using the even nodes as one cluster of individuals and the odd nodes as a subcluster.

c. Input Parameters

[0054] The methodology used in retrieval calculations **120**—computing the stochastic measurements—was based on: 1) generating the canonical representation of the raw multimedia documents—an application-specific document of structural information, and 2) applying the stochastic optimization retrieval algorithms to determine NNCs **190**—computing the raw fitness, standardized fitness, and adjusted fitness.

d. Synchronization Points

[0055] FIG. **1** provides periodic synchronization points **165**,**175**,**180** used for consistency restoration. Using a self-scheduling policy, the load-balancing model distributes the multimedia documents **105** that comprise the document dataset for each iteration. This random approach to the distribution of documents enables the system to adapt to each machine's characteristics at various stages of this iterative process **100-180**. By requiring that each node start each iteration **100**,**110**,**125** on the basis of a consistent state, the synchronization points are used to restore a consistent global state. FIG. **1** allows for continuous updates and redistribution

of multimedia documents **105,115,160,170** which incorporate the local and system-wide computational parameter adjustments.

[0056] The need for synchronization points **165,175,180** can be traced to scientific applications that are known to exhibit a diverse set of I/O access patterns. These are known as:

[0057] 1. Compulsory

[0058] 2. Checkpoint/restart

[0059] 3. Regular snapshots of the computation's progress

[0060] 4. Out-of-core read/writes

[0061] 5. Continuous output of data for visualization and other post-processing

The variability in the canonical document size accounts for the seemingly high random file accesses. Combining the file access patterns of all the indexers in the system reflects their compulsory nature. The synchronization points **165,175,180** provide the I/O checkpoints. The regular snapshots of the computation's progress are reflected in the intermediate solutions **160,170** that are created at the end of each iteration **165,175,180**.

e. Local Optimizing of the Object Parameter Sets

[0062] FIG. **2** lists the steps for illustrates a flow diagram for optimizing object parameters of parameter sets using assessment means, scoring means, stochastic means, and organizing means to generating ranking scores, indexing, scores, and clustering scores to obtain an ever increasing and better understanding of an ever-changing environment in order to find various types of hidden knowledge based on the complexity of the implemented model.

[0063] The discovery of hidden knowledge for ranking, indexing, and clustering multimedia documents incorporates two broadly labeled groups known as summarization algorithms—which find concise information contained in the input data, and anomaly detection algorithms—which identify unusual features of the data, to optimize the object parameters of parameter set. A subgrouping within these approaches stems from classificatory algorithms which partition data into disjoint groups.

[0064] The major components associated with the optimization model:

[0065] 1) Assement means to generate object parameters to represent problems specific discrete attributes

[0066] 2) Scoring means to provide a numerical assessment of the quality of discrete problem attributes which are represented as object parameters

[0067] 3) Stochastic means to provide classification tasks, knowledge engineering tasks, and generating hyperplane partitions within the search space

[0068] 4) Organizing means to provide historical aspects of the object parameter sets (preservation of indexing, ranking, and clustering scores for each search query set term)

[0069] The major steps for this approach are shown below

[0070] 1) Select a node as the source of the search strings, and create probe set T **210**

[0071] 2) For each data attribute, collect ranking measures based on hyperplane composition **220**

[0072] 3) Compute the hyperplane ranking **230**

[0073] 4) Compute localized similarity measures **240**

[0074] 5) Optimizing object parameters of parameter sets using assessment means, scoring means, stochastic means, and organizing means **250**

[0075] 6) Determine the members of the distinct species (nearest neighbors) **260**

[0076] 7) Apply the selection and recombination operators to evolve a new generation **270**

[0077] 8) Repeat above steps (1-6) **200-280**

f. Optimizing the Parameter Sets

[0078] The global optimization of parameter sets is guided scoring means and stochastic means leading to the intermediate state of the formulation of Knn **300** as shown in FIG. **3**. The members of each Knn are a result of applying scoring means to each parameter set to determine its rank among the population of parameter sets. The application of stochastic means to the members of each Knn leads to selective variations **320**, constructive variations **380**, clustering variations **340**, and stochastic variations **360**. The indexing, ranking, and clustering of parameter sets following formulation of Knn **300** is needed before applying stochastic means to determine nearest neighbors within each Knn using chosen object parameters. Applying stochastic means to each Knn clusters of parameter sets for chosen object parameters for selective variations **320** results in clustering variations **340** in the composition of the selected parameter sets. The clustering variations **340** can occur two or more times before selective variations **320**, constructive variations **380**, and stochastic variations **360** occurs. Likewise, the combination of clustering variations **340** and constructive variations **380**, or clustering variations **340** and stochastic variations **360** can occur zero or more times before selective variations **320** occurs following the formulation of Knn **300**. Following the formulation of Knn **300**, various combinations of selective variations **320**, constructive variations **380**, clustering variations **340**, and stochastic variations **360** can occur zero or more times. Scoring means are re-applied after the all variations are completed.

[0079] While particular embodiments and applications of the present invention have been illustrated and described herein, it is understood that the invention is not limited to the precise construction and components disclosed herein and that various modifications, changes, and variations may be made in the arrangement, operation, and details of the methods and apparatuses of the present invention without departing from the spirit and scope of the invention as it is defined in the appended claims.

I claim:

1. A method for indexing, ranking, and clustering multimedia documents by optimizing parameter sets of object parameters using assessment means, scoring means, stochastic means, and organizing means that guide selective variations, constructive variations, clustering variations, and stochastic variations comprising the steps of:

creating an initial population of a plurity of individual parameter sets based on the multimedia documents, the parameter sets comprising information sharing system object parameters for describing a model, structure, shape, design, process, search query set, and dynamic search space to optimize;

assessing the quality of a parameter set using the combination of two or more object parameters applying assessment means, scoring means, and stochastic means to guide search queries based on indexing score values, ranking score values, and clustering score values for each index term in the search query set object parameter;

optimizing a parameter set using the combination of two or more object parameters applying scoring means and

stochastic means to guide selective variations, constructive variations, clustering variations, and stochastic variations in nearest neighbor clusters of parameter sets;

formulating nearest neighbor clusters of parameter sets for the transmission of cultural information resulting from applying scoring means and stochastic means to two or more object parameters until no nearest neighbor clusters of two or more parameter sets are found;

grouping indexing scores, ranking scores, and clustering scores using organizing means document query terms of the search query set object parameter to form structure index term object parameters for each query term; and

repeating all steps until achieving a periodic optimal multimedia clusters for combinations two or more object parameters of parameter sets.

2. The method of claiml wherein the indexing, ranking, and clustering of the parameter sets using combinations of two or more object parameters guided by applying assessment means, scoring means, and stochastic means to the document query terms of the search query set object parameter.

3. The method of claim **2** wherein selective variations, constructive variations, clustering variations, and stochastic variations of parameter sets is guided by applying scoring means and stochastic means to the terms of the search query set object parameter.

4. The method of claim **2** wherein selective variations, constructive variations, clustering variations, and stochastic variations of parameter sets is guided by applying scoring means and stochastic means to two or more object parameters of parameter sets.

5. The method of claim **2** wherein selective variations, constructive variations, clustering variations, and stochastic variations of parameter sets is guided by applying scoring means to and stochastic means to parameter sets.

6. The method of claim **1** wherein formulation of nearest neighbor clusters of parameter sets is guided by applying scoring means and stochastic means to determine the rank of each parameter set.

7. The method of claim **6** wherein selective variations of parameter sets is guided by applying scoring means and stochastic means to a selected sets of object parameters in each parameter set.

8. The method of claim **7** wherein clustering variations of parameter sets is guided by applying scoring means and stochastic means to selected sets of object parameters in each parameter set.

9. The method of claim **8** wherein constructive variations of parameter sets is guided by applying scoring means and stochastic means to selected sets of object parameters in each parameter set.

10. The method of claim **8** wherein stochastic variations of parameter sets is guided by applying scoring means and stochastic means selected sets of object parameters in each parameter set.

11. A system that indexes, ranks, and clusters multimedia documents by optimizing parameter sets of object parameters using selective variations, constructive variations, clustering variations, and stochastic variations comprising the steps of:

creating an initial population of parameter sets based on the multimedia documents, the parameter sets comprising information sharing system object parameters for describing a model, structure, shape, design, process, search query set, and dynamic search space to optimized;

assessing the quality of each parameter set in nearest neighbor clusters of parameter sets by assessment means, scoring means, and stochastic means following selective variations, constructive variations, clustering variations, and stochastic variations;

ranking each parameter set in nearest neighbor clusters of parameter sets by scoring means and stochastic means following selective variations, constructive variations, clustering variations, and stochastic variations;

formulating clusters of parameter sets for the transmission of cultural information resulting from selective variations, constructive variations, clustering variations, and stochastic variations until no nearest neighbor clusters of two or more parameter sets are found;

improving stochastically selected object parameters of two or more parameter sets by selective variations, constructive variations, clustering variations, and stochastic variations, and

repeating all steps until achieving a periodic optimal multimedia document clusters parameter sets for all possible document query terms of the search query set object parameter.

12. A system of claim **11** wherein the indexing, ranking, and clustering of multimedia documents within the population of parameter sets is guided by selective variations, constructive variations, clustering variations, and stochastic variations.

13. A system of claim **11** wherein the quality of a parameter set within the population of parameter sets is stochastically improved by selective variations, constructive variations, clustering variations, and stochastic variations.

14. A system of claim **11** wherein the formulation of nearest neighbor clusters within the population of parameter sets is guided by using scoring means and stochastic means to rank parameter sets.

15. A system of claim **11** wherein all the object parameters within parameter sets within the population of parameter sets are stochastically improved by selective variations, constructive variations, clustering variations, and stochastic variations leads to stochastic improvements of parameter sets.

* * * * *