REAL TIME DATA WAREHOUSING

DESCRIPTION

CROSS-REFERENCE TO RELATED APPLICATIONS

The present application claims the benefit of provisional application number 60/344,067, filed in the United States Patent Office on December 28, 2001.

FEDERALLY SPONSORED OR DEVELOPMENT

Not Applicable.

TECHNICAL FIELD:

This invention generally relates to a method, program and system for processing and retrieving data in a data warehouse and, more particularly, to a method, program and system for the processing of data into and in a data warehouse, to the querying of data in a data warehouse, and the analyzing of data in a data warehouse.

BACKGROUND OF THE INVENTION:

Data warehouses are computer-based databases designed to store records and respond to queries generally from multiple sources. The records correspond with entities, such as individuals, organizations and property. Each record contains identifiers of the entity, such as for example, a name, address or account information for an individual.

Unfortunately, the effectiveness of current data warehouse systems is diminished because of certain limitations that create, perpetuate and/or increase certain data quality, integrity and performance issues. Such limitations also increase the risk, cost and time required to implement, correct and maintain such systems.

The issues and limitations include, without limitation, the following: (a) challenges associated with differing or conflicting formats emanating from the various sources of data, (b) incomplete data based upon missing information upon receipt, (c) multiple records entered that reflect the same entity based upon (often minor) discrepancies or misspellings, (d) insufficient capability to identify whether multiple records are reflecting the same entity and/or whether there is some relationship between multiple records, (e) lost data when two

records determined to reflect the same entity are merged or one record is discarded, (f) insufficient capability to later separate records when merged records are later determined to reflect two separate entities, (g) insufficient capability to issue alerts based upon user-defined alert rules in real-time, (h) inadequate results from queries that utilize different algorithms or conversion processes than the algorithms or conversion processes used to process received data, and (i) inability to maintain a persistent query in accordance with a pre-determined criteria, such as for a certain period of time.

For example, when the identifiers of an individual are received and stored in a database: (a) the records from one source may be available in a comma delimited format while the records of another source may be received in another data format; (b) data from various records may be missing, such as a telephone number, an address or some other identifying information; or (c) two records reflecting the same individual may be unknowingly received because one record corresponds to a current name and another record corresponds to a maiden name. In the latter situation, the system may determine that the two records ought to be merged or that one record (perhaps emanating from a less reliable source) be discarded. However, in the merging process, current systems typically abandon data, which negates the ability to later separate the two records if the records are determined to reflect two separate entities.

Additionally, when the identifiers are received and stored in a database, the computer may perform transformation and enhancement processes prior to loading the data into the database. However, the query tools of current systems use few, if any, of the transformation and enhancement processes used to receive and process the received data, causing any results of such queries to be inconsistent, and therefore inadequate, insufficient and potentially false.

Similarly, current data warehousing systems do not have the necessary tools to fully identify the relationship between entities, or determine whether or not such entities reflect the same entity in real-time. For example, one individual may have the same address of a second individual and the second individual may have the same telephone number of a third individual. In such circumstances, it would be beneficial to determine the likelihood that the first individual had some relationship with the third individual, especially in real-time.

Furthermore, current data warehousing systems have limited ability to identify inappropriate or conflicting relations between entities and provide alerts in real-time based

upon user-defined alert rules. Such limited ability is based upon several factors, including, without limitation, the inability to efficiently identify relationships as indicated above.

Furthermore, current data warehousing systems cannot first transform and enhance a record and then maintain a persistent query over a predetermined period. A persistent query would be beneficial in various circumstances, including, without limitation, in cases where the name of a person is identified in a criminal investigation. A query to identify any matches corresponding with the person may initially turn up with no results and the queried data in current systems is essentially discarded. However, it would be beneficial to load the query in the same way as received data wherein the queried data may be used to match against other received data or queries and provide a better basis for results.

As such, any or all the issues and limitations (whether identified herein or not) of current data warehouse systems diminishes accuracy, reliability and timeliness of the data warehouse and dramatically impedes performance. Indeed, the utilization with such issues may cause inadequate results and incorrect decisions based upon such results.

The present invention is provided to address these and other issues.

SUMMARY OF THE INVENTION:

It is an object of the invention to provide a method, program and system for processing data into and in a database. The method preferably comprises the steps of: (a) receiving data for a plurality of entities, (b) utilizing an algorithm to process the received data, (c) storing the processed data in the database, (d) receiving data queries for retrieving data stored in the database, and (e) utilizing the same algorithms to process the queries.

The data comprises one or more records having one or more identifiers representing one or more entities. The entities may be individuals, property, organizations, proteins or other things that can be represented by identifying data.

The algorithm includes receiving data that has been converted to a standardized message format and retains attribution of the identifiers, such as a source system, the source system's unique value for the identifier, query system and/or user.

The algorithm process includes analyzing the data prior to storage or query in the database wherein such analyzing step may include: (a) comparing one or more identifiers against a user-defined criterion or one or more data sets in a database, list, or other electronic format, (b) formatting the identifier in accordance with the user-defined standard, (c)

enhancing the data prior to storage or query by querying one or more data sets in other databases (which may have the same algorithm as the first database and continue to search in a cascading manner) or lists for additional identifiers to supplement the received data with any additional identifiers, (d) creating hash keys for the identifiers, and (d) storing processed queries based upon user-defined criterion, such as a specified period of time.

It is further contemplated that the method, program and system would include: (a) utilizing an algorithm to process data and match records wherein the algorithm process would: (i) retrieve from the database a group of records including identifiers similar to the identifiers in the received data, (ii) analyze the retrieved group of records for a match to the received data, (iii) match the received data with the retrieved records that are determined to reflect the same entity, (iv) analyze whether any new identifiers were added to any matched record, and (v) re-search the other records of the retrieved group of records to match to any matched record, and (b) storing the matched records in the database. Additionally, the algorithm may include: (a) retrieving from the database an additional group of records including identifiers similar to the identifiers in the matched record, (b) repeating the steps of retrieving records, analyzing for matches, matching same entity records, analyzing new identifiers, and re-searching retrieved records until no additional matches are found, and (c) assigning a persistent key to the records. Such processes could be performed in batch or in real-time.

It is yet further contemplated that the method, program and system includes determining whether a particular identifier is common across entities or generally distinctive to an entity, and separating previously matched records if the particular identifier used to match the records is later determined to be common across entities and not generally distinctive of an entity. Such determining and separating steps may be performed in real-time or in batch. The determining and separating steps may include stopping any additional matches based upon an identifier that is determined to be common across entities and not generally distinctive of an entity, as well as re-processing any separated records.

It is further contemplated that the received data is compared with at least one other previously stored record to determine the existence of a relationship between the entities, and that a relationship record is created for every two entities for which there exists a relationship. The relationship record may include confidence indicator(s), indicating the likelihood of a relationship between the two entities or the likelihood that the two entities are

the same. The relationship record may also reference roles of the entities that are included in the received data or assigned. The relationship records are analyzed to determine the existence of any previously unknown related records based upon the existence of a user-defined criterion. The relationship records reflect a first degree of separation which may be analyzed and navigated to include only those records that meet a predetermined criterion, such as a maximum number of degrees of separation test or a minimum level of the relationship and/or likeness confidence indicators. An alert may be issued identifying the group of related records based upon a user-defined alert rule. The alert may be communicated through various electronic communication means, such as an electronic mail message, a telephone call, a personal digital assistant, or a beeper message.

It is further contemplated that the method would include: (a) duplicating the relationship records on one or more databases, (b) distributing received data to one or more of the additional databases for analysis based upon work load criteria; and (c) issuing any alerts from the additional databases.

It is further contemplated that the method and system would include transferring the stored data to another database that uses the same algorithm as the first database. The steps of processing and transferring may be performed in real-time or in batch.

These and other aspects and attributes of the present invention will be discussed with reference to the following drawings and accompanying specification.

BRIEF DESCRIPTION OF THE DRAWINGS:

FIGURE 1 is a block diagram of a system in accordance with the present invention;

FIGURE 2 is a flow chart for process data in the System block in FIGURE 1;

FIGURES 3A-3C are a flow chart of the Process Algorithm block in FIGURE 2; and

FIGURES 4A-4B are a flow chart of the Evaluate Stored Analyzed Record block in FIGURE 3.

DETAILED DESCRIPTION OF THE INVENTION:

While this invention is susceptible of embodiment in many different forms, there is shown in the drawing, and will be described herein in detail, specific embodiments thereof with the understanding that the present disclosure is to be considered as an exemplification of the principles of the invention and is not intended to limit the invention to the specific

embodiments illustrated.

A data processing system 10 for processing data into and in a database and for retrieving the processed data is illustrated in Figures 1-4B. The system 10 includes at least one conventional computer 12 having a processor 14 and memory 16. The memory 16 is used for storage of the executable software to operate the system 10 as well as for storage of the data in a database and random access memory. However, the software can be stored or provided on any other computer readable medium, such as a CD, DVD or floppy disc. The computer 12 may receive inputs from a plurality of sources 18_1 - 18_n.

The data comprises one or more records having one or more identifiers representing one or more entities. The entities may be individuals, organizations, property, proteins, chemical or organic compounds, biometric or atomic structures, or other things that can be represented by identifying data. The identifiers for an individual type entity may include the individual's name, address(es), telephone number(s), credit card number(s), social security number, employment information, frequent flyer or other loyalty program, or account information. Generally distinctive identifiers are those that are distinctive to a specific entity, such as a social security number for an individual entity.

The system 10 receives the data from the plurality of sources 18_1 - 18_n and utilizes an algorithm 22 to process the received data 20. The algorithm is stored in the memory 16 and is processed or implemented by the processor 14.

The received data 20 including, without limitation, attributions of the received data (e.g., source system identification), is likely received in many data formats. Prior to being processed by the algorithm 22, the received data 20 is converted into a standardized message format 24, such as Universal Message Format.

Thereafter, as illustrated in FIGURES 3A-3C, the algorithm 22 receives the standardized data 26 and analyzes 28 the received data 26 prior to storage or query in the database by: (a) comparing the received data 26 to user-defined criteria or rules to perform several functions, including, without limitation, the following: (i) name standardization 30 (e.g., comparing to a root names list), (ii) address hygiene 32 (e.g., comparing to postal delivery codes), (iii) field testing or transformations 34 (e.g., comparing the gender field to confirm M/F or transforming Male to M, etc.), (iv) user-defined formatting 36 (e.g., formatting all social security numbers in a 999-99-9999 format), (b) enhancing the data 38 by causing the system 10 to access one or more databases 40 (which may contain the same

algorithm as the first database, thus causing the system to access additional databases in a cascading manner) to search for additional information (which may be submitted as received data 20) which can supplement 42 the received data 26, and (c) building hash keys of the analyzed data 44. Any new, modified or enhanced data can be stored in newly created fields to maintain the integrity of the original data. For example, if the name "Bobby Smith" is received in a standardized format 26, the name "Bobby" may be compared to a root name list 30, standardized to the name "Robert" and saved in a newly created field for the standard name. Additionally, if the name and address for Bobby Smith is received 26, the system 10 can access a conventional Internet-based people finder database 40 to obtain Bobby Smith's telephone number, which can then be formatted in a standard way based upon user-defined criteria 36. Furthermore, the address field may be compared to an address list 32, resulting in the text "Street" added to the end of the standardized address. Hash keys are then built 44 based upon the enhanced data and stored in newly created fields.

The system 10 also receives queries 46 from the plurality of sources 18_1 - 18_n and utilizes the same algorithm 22 to analyze and process the received queries 46. For example, if a query for "Bobby Smith" is received 46, the same algorithm 22 which standardized the received name "Bobby" to the name "Robert" will also standardize the queried name "Bobby" to the queried name "Robert." Indeed, the system 10 loads and stores received queries 46 the same as received data 20, maintaining the full attribution of the query system and user. As such, as the system 10 processes the received queries 46, the algorithm 22 may search other databases 40, such as a public records database, to find missing information. Query results 94 may be broader than exact matches, and may include relationship matches. For example, if the query is for "Bobby Smith", the query results 94 may include records of people who have used Bobby Smith's credit card, or have lived at Bobby Smith's address.

The algorithm 22 also performs a function upon receipt of any received data 26 to: (a) determine whether there is an existing record in the database that matches the entity corresponding to such received data and (b) if so, matching the received data to the existing record. For example, the algorithm retrieves a group of records 48 (including identifiers similar to the identifiers in the received data) from the database for possible candidates and analyzes the retrieved group of records for a match 50 identifying an existing stored record corresponding to the received data based upon generally distinctive identifiers 52. If a match is identified 54, the algorithm analyzes whether the matched record contains any new or

previously unknown identifiers 56. If there were new or previously unknown identifiers 56, the algorithm 22 would analyze the new or previously unknown identifiers 58, add or update the candidate list/relationship records 70 based upon the new or previously unknown identifiers in the matched record, and determine whether any additional matches 50 exist. This process is repeated until no further matches can be discerned. The matching process would then assign all of the matched records the same persistent key 60. Furthermore, if no matches were found for any record, the unmatched record would be assigned its own persistent key 62. The records retain full attribution of the data and the matching process does not lose any data through a merge, purge or delete function.

For example, if record #1 has an individual's name, telephone number and address, and record #2 has the same name and a credit card number. One does not know whether or not they are the same individual, so the records must be kept separate. Then data for record #3 is received, including the individual's name (same as record #1), address (same as record #1), telephone number (same as record #1) and credit card number. Because the name, telephone number and address for #1 and #3 match, the system 10 may determine that #1 and #3 are describing the same individual, so the algorithm matches record #1 with #3 data. The system 10 then re-runs the algorithm, comparing the matched record #1 with the other records of the candidate list or additional records that include identifiers similar to the matched record. Because the name and credit card number of matched record #1 matches the name and credit card number of record #2, these two records are also matched. This matched record is then run again against the candidate list or additional records retrieved looking for matches 54 until no more matches are obtained.

On occasion, the system 10 may determine that two records were incorrectly matched. For example, social security numbers are considered generally distinctive identifiers for individuals, and thus records often are matched based upon the same social security number. However, it is possible that such number, in certain circumstances, is later determined to be common across entities and not generally distinctive of an entity. For example, consider a data entry operation having a record field for social security numbers as a required field, but the data entry operator who did not know the social security number of the individuals merely entered the number "123-45-6789" for each individual.

In such a case, the social security number would be common across such individual type entities and no longer a generally distinctive identifier for these individuals.

Accordingly: (a) the now known common identifier would be added to a list of common identifiers and all future processes would not attempt to retrieve records for the candidate list or create relationship records 70 based upon the now known common identifier, thus stopping any future matches 64 and (b) any records that were matched based upon that erroneous social security number would need to be split to reflect the data prior to the match, thus requiring no prior data loss. To accomplish the latter objective, the system 10 separates any matches that occurred based upon the incorrect assumptions 66 to the point prior to the incorrect assumption pursuant to the full attribution of the data, without any loss of data. Thus, if record #1 for "Bobby Smith" (which had been standardized to "Robert Smith") had been matched with record #2 for "Robert Smith", and it is later determined that these are two different individuals, and that they needed to be broken into the original record #'s 1 and 2, the algorithm would identify that the standardized "Robert Smith" of record #1 was known as "Bobby." Furthermore, the determining and separating steps can be performed in real-time or in batch. Furthermore, the separated records may be re-submitted as new received data to be processed in the system.

There are also times when relationships, even less than obvious relationships, need to be evaluated 68. For example, individuals #1 and #2 may each have a relationship to an organization #3. Thus it is possible, perhaps likely, that there is a relationship between individuals #1 and #2. The relationships can be extended to several degrees of separation. Accordingly, the system 10 compares all received data to all records in the stored data and creates a relationship record 70 for every pair of records for which there is some relationship between the respective entities. The relationship record 70 would include relationship types (e.g., father, co-conspirator), the confidence indicators (which are scores indicating the strength of relationship of the two entities) 72 and the assigned persistent key 60 or 62. For example, the confidence indicators 72 may include a relationship score and a likeness score. The relationship score is an indicator, such as between 1 and 10, representing the likelihood that there is a relationship between individual #1 and individual #2. The likeness score is also an indicator, such as between 1 and 10, that individual #1 is the same person as individual #2. The confidence indicators 72 could be identified during the matching process described hereinabove.

The system 10 also analyzes the received data 20 and queries 46 to determine the existence of a condition that meets the criteria of a user-defined alert rule 74, such as an

inappropriate relationship between two entities or a certain pattern of activities based upon relationship records that have a confidence indicator greater than a predetermined value and/or have a relationship record less than a predetermined number of degrees of separation. For example, the system 10 may include a list of fraudulent credit cards that could be used to determine whether any received data or query contains a credit card number that is on the list of fraudulent credit card numbers. Additionally, the user-defined alert rule 74 may cause the received data and queries to be reported. For example, an alert rule may exist if, upon entering data of a new vendor, it was determined that the new vendor had the same address as a current employee, indicating a relationship between the vendor and the employee that perhaps the employer would like to investigate. Upon determination of a situation that would trigger the user-defined alert rule, the system 10 issues an alert 74 which may be communicated through various mediums, such as a message via an e-mail or to a hand-held communication device, such as an alpha-numeric beeper, personal digital assistant or a telephone.

For example, based upon a user-defined alert rule for all records that have a likelihood of relationship confidence indicator greater than seven 76 to a maximum of six degrees of separation 78, the system 10 will: (a) start with individual #1, (b) find all other individuals 80 related to #1 having a confidence indicator greater than seven 76, (c) analyze all of the first degree of separation individuals 80, and determine all individuals 82 related to the first degree of separation individuals 80 having a confidence indicator greater than seven 84 and (d) repeat the process until it meets the six degrees of separation parameter 78. The system would send electronically an alert 74 (that may include all the resulting records based upon a user-defined criterion) to the relevant individual or separate system enabling further action.

Furthermore, the relationship records 70 could be duplicated over several databases. Upon receipt of received data 20, the system could systematically evaluate the nature of the work load of each of the other databases and distribute the matched/related/analyzed records to the database most likely to efficiently analyze the stored analyzed record 68. Any alerts 74 could then be issued from any results emanating from the other databases.

Finally, the processed data can be transferred 88 to additional databases based upon a cascading warehouse publication list 86 that may utilize the same algorithm 92, either on a real-time or batch process. In this manner, the transferred data 88 can then be used to match

with data (which may include different data) in the additional databases and any subsequent database to identify relationships, matches or processing of such data. For example, the matched records based upon the confidence indicators in a local database may be transferred 88 to the regional database to be compared and matched with data utilizing the same algorithm 92. Thereafter, the processed data resulting from the regional database may be transferred 88 to the national office. By combining the processed data in each step, especially in real-time, organizations or system users would be able to determine inappropriate or conflicting data prompting further action.

Conventional software code can be used to implement the functional aspects of the method, program and system described above. The code can be placed on any computer readable medium for use by a single computer or a distributed network of computers, such as the Internet.

From the foregoing, it will be observed that numerous variations and modifications may be effected without departing from the spirit and scope of the invention. It is to be understood that no limitation with respect to the specific apparatus illustrated herein is intended or should be inferred. It is, of course, intended to cover by the appended claims all such modifications as fall within the scope of the claims.

CLAIMS

WHAT IS CLAIMED IS:

1. A method for processing data comprising the steps of:

receiving data comprising at least one record having at least one identifier, each record representing at least one of a plurality of entities;

utilizing an algorithm to process the received data;

storing the processed data in a database;

receiving data queries for retrieving at least a portion of the data stored in the database; and

utilizing the algorithm to process the queries.

2. The method of claim 1 wherein the entities are people.

3. The method of claim 1 wherein the entities are personal property.

4. The method of claim 3 wherein the personal property are vehicles.

5. The method of claim 1 wherein the entities are real property.

6. The method of claim 1 wherein the entities are organizations.

7. The method of claim 1 wherein the entities are chemical compounds.

8. The method of claim 1 wherein the entities are organic compounds.

9. The method of claim 1 wherein the entities are proteins.

10. The method of claim 1 wherein the entities are biological structures.

11. The method of claim 1 wherein the entities are biometric values.

12. The method of claim 1 wherein the entities are atomic structures.

13. The method of claim 1 further comprising the step of converting the received data into a standardized message format prior to utilizing an algorithm to process the received data.

14. The method of claim 1 wherein the step of utilizing an algorithm to process the received data includes retaining an attribution of each record.

15. The method of claim 14 wherein the step of retaining an attribution of each record includes retaining an identity of:

a source system providing each record and

a unique identifier representing each record in the source system.

16. The method of claim 14 wherein the step of retaining an attribution of each record includes retaining an identity of a query system and a particular user.

17. The method of claim 1 wherein the step of utilizing an algorithm to process the received data includes analyzing the received data prior to one of storage in the database and query in the database.

18. The method of claim 17 wherein the step of analyzing the received data prior to one of storage in the database and query in the database includes comparing at least one of the identifiers against one of:

a user-defined criterion and

at least one data set in one of a secondary database and a list.

19. The method of claim 18 wherein the compared identifier is a name of at least one of the plurality of entities and the data set is in a names root list.

20. The method of claim 18 wherein the compared identifier is an address of at least one of the plurality of entities and the data set is in an address list.

21. The method of claim 18 wherein the step of comparing at least one of the identifiers against a user-defined criterion includes formatting at least one identifier in accordance with the user-defined standard.

22. The method of claim 18 wherein the step of analyzing the received data prior to one of storage in the database or query in the database includes enhancing the received data.

23. The method of claim 22 wherein the step of enhancing the received data includes:

querying at least one data set in one of the secondary database and the list for additional identifiers for the received data, and

supplementing the received data with the additional identifiers.

24. The method of claim 23 wherein the step of querying at least one data set includes:

at least one data set being in the secondary database utilizing the algorithm to query additional databases to locate additional identifiers relating to at least one of the received identifiers; and

supplementing the received data with the additional identifiers located in the secondary database.

25. The method of claim 17 wherein the step of analyzing the received data prior to one of storage in the database and query in the database includes creating hash keys of the identifiers.

26. The method of claim 1 wherein the step of utilizing an algorithm to process received

data includes storing in the database processed queries based upon a user-defined criterion.

27. The method of claim 26 wherein the user-defined criterion includes an expiration date.

28. The method of claim 1 wherein the steps of receiving data comprising at least one record having at least one identifier, each record representing at least one of a plurality of entities, utilizing an algorithm to process the received data, and storing the processed data in a database are performed in real-time.

29. The method of claim 1 wherein the steps of receiving data comprising at least one record having at least one identifier, each record representing at least one of a plurality of entities, utilizing an algorithm to process the received data, and storing the processed data in a database are performed in batch.

30. The method of claims 1 or 17, wherein the step of utilizing an algorithm to process the received data includes:

retrieving from the database a group of additional records having identifiers similar to the identifiers in the received data;

analyzing each identifier of the retrieved group of records for a match to at least a portion of the received data;

matching at least a portion of the received data with at least one analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities;

analyzing whether at least one identifier is included in the at least a portion of the received data that was not previously stored in the at least one analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities; and

re-analyzing each identifier of the retrieved group of records for a match to:

at least a portion of the received data and

the analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities; and

storing the matched records in the database.

31. The method of claim 30 wherein matching at least a portion of the received data with

at least one analyzed record includes assigning a persistent key.

32. The method of claim 30 wherein the step of utilizing an algorithm to process the received data further comprises retrieving from the database an additional group of records having identifiers similar to the identifiers in:

at least a portion of the received data and

the analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities prior to re-analyzing each identifier of the retrieved group of records for a match.

33. The method of claim 32 wherein utilizing an algorithm to process the received data includes repeating:

retrieving from the database a group of records;

analyzing each identifier of the retrieved group of records;

matching at least a portion of the received data;

analyzing whether at least one identifier is included in the at least a portion of the received data that was not previously stored;

retrieving from the database an additional group of records; and

re-analyzing each identifier of the retrieved group of records for a match until no additional matches are determined.

34. The method of claim 30 wherein the step of utilizing an algorithm to process the received data includes:

determining whether a particular identifier is one of:

common across records representing at least two different entities and

generally distinctive of a record representing a particular entity;

and

separating records that were previously matched based on a particular identifier if the particular identifier is determined to be common across records representing at least two different entities and not generally distinctive of a record representing a particular entity.

35. The method of claim 34 wherein the step of utilizing an algorithm to process the received data includes prohibiting any additional matches of records based on a particular identifier if the particular identifier is determined to be common across records representing at least two different entities and not generally distinctive of a record representing a particular entity.

at least one analyzed record includes assigning a persistent key.

32. The method of claim 30 wherein the step of utilizing an algorithm to process the received data further comprises retrieving from the database an additional group of records having identifiers similar to the identifiers in:

at least a portion of the received data and

the analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities

prior to re-analyzing each identifier of the retrieved group of records for a match.

33. The method of claim 32 wherein utilizing an algorithm to process the received data includes repeating:

retrieving from the database a group of records;

analyzing each identifier of the retrieved group of records;

matching at least a portion of the received data;

analyzing whether at least one identifier is included in the at least a portion of the received data that was not previously stored;

retrieving from the database an additional group of records; and re-analyzing each identifier of the retrieved group of records for a match until no additional matches are determined.

34. The method of claim 30 wherein the step of utilizing an algorithm to process the received data includes:

determining whether a particular identifier is one of:

common across records representing at least two different entities and

generally distinctive of a record representing a particular entity;

and

separating records that were previously matched based on a particular identifier if the particular identifier is determined after a first occurrence to be common across records representing at least two different entities and not generally distinctive of a record representing a particular entity.

35. The method of claim 34 wherein the step of utilizing an algorithm to process the received data includes prohibiting any additional matches of records based on a particular identifier if the particular identifier is determined to be common across records representing at least two different entities and not generally distinctive of a record representing a particular entity.

an entity represented by the particular record having a relationship with the portion of the received data, and

an entity represented by the portion of the received data are the same.

45. The method of claim 40 wherein the step of utilizing an algorithm to process received data includes analyzing the relationship records to determine whether the relationship records reflect at least one relationship not previously determined.

46. The method of claim 45 wherein the step of analyzing the relationship records includes analyzing relationship records reflecting at least one level of degrees of separation.

47. The method of claim 46 wherein the step of analyzing relationship records reflecting at least one level of degrees of separation includes analyzing relationship records meeting at least one user-defined criterion.

48. The method of claim 47 wherein the step of analyzing relationship records meeting at least one user-defined criterion includes limiting the relationship records analyzed to a maximum level of degrees of separation.

49. The method of claim 47 wherein the step of analyzing relationship records meeting at least one user-defined criterion includes limiting the relationship records analyzed to relationship records that include confidence indicators greater than a minimum amount.

50. The method of claim 45 wherein the step of utilizing an algorithm to process received data further comprises issuing an alert based upon at least one user-defined alert rule.

51. The method of claim 50 wherein the step of issuing an alert based upon at least one user-defined alert rule includes having the alert communicated via electronic communications means.

52. The method of claim 51 wherein the electronic communications means comprise an e-mail system.

53. The method of claim 51 wherein the electronic communications means comprise a telephone.

54. The method of claim 51 wherein the electronic communications means comprise a beeper.

55. The method of claim 51 wherein the electronic communications means comprise a

personal digital assistant.

56. The method of claim 50 wherein the step of analyzing the relationship records includes:

duplicating the relationship records on at least one secondary database;

distributing received data to the at least one secondary database for analysis based upon a work load criteria; and

issuing the alert meeting the criteria of a user-defined alert rule from the at least one secondary database.

57. The method of claims 1 or 28 wherein the step of utilizing an algorithm to process the received data further comprises transferring the stored processed data to at least one secondary database utilizing the algorithm.

58. The method of claim 57 wherein the step of transferring the stored processed data to at least one secondary database is performed in real-time.

59. The method of claim 57 wherein the step of transferring the stored processed data to at least one secondary database is performed in batch.

60. A method for processing data comprising the steps of:

receiving data comprising at least one record having at least one identifier, each record representing at least one of a plurality of entities;

utilizing an algorithm to:

retrieve from a database a group of additional records having identifiers similar to the identifiers in the received data,

analyze each identifier of the retrieved group of records for a match to at least a portion of the received data,

match the at least a portion of the received data with at least one analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities,

analyze whether at least one identifier is included in the at least a portion of the received data that was not previously stored in the at least one analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities; and

re-analyze each identifier of the retrieved group of records for a match to:

the at least a portion of the received data and

the analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities; and

storing the matched records in the database.

61. The method of claim 60 wherein the step of utilizing an algorithm includes assigning a persistent key.

62. The method of claim 60 wherein the step of utilizing an algorithm further comprises retrieving from the database an additional group of records having identifiers similar to the identifiers in:

at least a portion of the received data and

the analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities prior to re-analyzing each identifier of the retrieved group of records for a match.

63. The method of claims 60 or 62 wherein the step of utilizing an algorithm includes repeating:

retrieving from a database a group of additional records;

analyzing each identifier of the retrieved group of records;

matching at least a portion of the received data;

analyzing whether at least one identifier is included in the at least a portion of the received data that was not previously stored;

retrieving from the database an additional group of records; and

re-analyzing each identifier of the retrieved group of records for a match until no additional matches are determined.

64. The method of claim 63 wherein the steps of receiving data, utilizing an algorithm and storing the matched records are performed in real-time.

65. The method of claim 63 wherein the steps of receiving data, utilizing an algorithm and storing the matched records are performed in batch.

66. The method of claim 60 wherein the step of utilizing an algorithm includes:

determining whether a particular identifier is one of:

common across records representing at least two different entities and

generally distinctive of a record representing a particular entity; and

separating records that were previously matched based on a particular identifier if the

particular identifier is determined to be common across records representing at least two different entities and not generally distinctive of a record representing a particular entity.

67. The method of claim 66 wherein the step of utilizing an algorithm includes prohibiting any additional matches of records based on a particular identifier if the particular identifier is determined to be common across records representing at least two different entities and not generally distinctive of a particular entity.

68. The method of claim 66 wherein the step of utilizing an algorithm includes re-processing the separated records as received data.

69. The method of claim 66 wherein the steps of determining whether a particular identifier is one of common across records representing at least two entities and generally distinctive of a record representing a particular entity and separating records that were previously matched are performed in real-time.

70. The method of claim 66 wherein the steps of determining whether a particular identifier is one of common across records representing at least two different entities and generally distinctive of a record representing a particular entity and separating records that were previously matched are performed in batch.

71. The method of claim 60 wherein the step of utilizing an algorithm includes:

comparing the received data with at least one stored record to determine the existence of a relationship; and

creating a relationship record for each stored record determined to reflect a relationship with at least a portion of the received data.

72. The method of claim 71 wherein the step of utilizing an algorithm includes creating at least one confidence indicator for each relationship record.

73. The method of claim 72 wherein the steps of comparing the received data, creating a relationship record, and creating at least one confidence indicator are performed in real-time.

74. The method of claim 72 wherein the steps of comparing the received data, creating a relationship record, and creating at least one confidence indicator are performed in batch.

75. The method of claim 72 wherein at least one of the confidence indicators indicates the likelihood of a relationship between:

an entity represented by the particular record having a relationship with the portion of

the received data, and

an entity represented by the portion of the received data.

76. The method of claim 72 wherein at least one of the confidence indicators indicates the likelihood that:

an entity represented by the particular record having a relationship with the portion of the received data, and

an entity represented by the portion of the receive data are the same.

77. The method of claim 72 wherein the step of utilizing an algorithm includes analyzing the relationship records to determine whether the relationship records reflect at least one relationship not previously determined.

78. The method of claim 77 wherein the step of analyzing the relationship records includes analyzing relationship records reflecting at least one level of degree of separation.

79. The method of claim 78 wherein the step of analyzing relationship records reflecting at least one level of degree of separation includes analyzing relationship records meeting at least one user-defined criterion.

80. The method of claim 79 wherein the step of analyzing relationship records meeting a user-defined criterion includes limiting the relationship records analyzed to a maximum level of degrees of separation.

81. The method of claim 79 wherein the step of analyzing relationship records meeting a user-defined criterion includes limiting the relationship records analyzed to relationship records that include confidence indicators greater than a minimum amount.

82. The method of claim 77 wherein the step of utilizing an algorithm further comprises issuing an alert based upon at least one user-defined alert rule.

83. The method of claim 82 wherein the step of issuing an alert based upon at least one user-defined alert rule includes having the alert communicated via electronic communications means.

84. The method of claim 83 wherein the electronic communications means comprise an e-mail system.

85. The method of claim 83 wherein the electronic communications means comprise a telephone.

86. The method of claim 83 wherein the electronic communications means comprise a beeper.

87. The method of claim 83 wherein the electronic communications means comprise a personal digital assistant.

88. The method of claim 82 wherein the step of analyzing the relationship records includes:

duplicating the relationship records on at least one secondary database;

distributing received data to the at least one secondary for analysis based upon a work load criteria; and

issuing the alert based upon the user-defined alert rule from the at least one secondary database.

89. The method of claim 60 further comprising the step of converting the received data into a standardized message format prior to the step of utilizing an algorithm.

90. The method of claim 60 wherein the step of utilizing an algorithm includes retaining an attribution of each of the identifiers.

91. The method of claim 90 wherein the step of retaining the attribution of each record includes retaining an identity of:

a source system providing each record and

a unique identifier representing each record in the source system.

92. The method of claim 90 wherein the step of retaining an attribution of each of the identifiers includes retaining an identity of a query system and a particular user.

93. The method of claim 60 wherein the step of utilizing an algorithm includes analyzing the received data prior to one of storage in the database and query in the database.

94. The method of claim 93 wherein the step of analyzing the received data prior to one of storage in the database and query in the database includes comparing at least one of the identifiers against one of:

a user-defined criterion and

at least one data set in one of a secondary database and a list.

95. The method of claim 94 wherein the compared identifier is a name of at least one of the plurality of entities and the data set is in a names root list.

96. The method of claim 94 wherein the compared identifier is an address of at least one of the plurality of entities and the data set is in an address list.

97. The method of claim 94 wherein the step of comparing at least one of the identifiers against a user-defined criterion includes formatting at least one identifier in accordance with the user-defined criterion.

98. The method of claim 93 wherein the step of analyzing the received data prior to one of storage in the database and query in a database includes enhancing the received data.

99. The method of claim 98 wherein the step of enhancing the received data includes:

querying at least one data set in one of the secondary database and the list for additional identifiers for the received data; and

supplementing the received data with the additional identifiers.

100. The method of claim 99 wherein the step of querying at least one data set includes:

at least one data set being in the secondary database utilizing the algorithm to query additional databases to locate additional identifiers relating to at least one of the received identifiers; and

supplementing the received data with the additional identifiers located in the secondary database.

101. The method of claim 93 wherein the step of utilizing an algorithm includes creating hash keys of the identifiers.

102. The method of claim 60 wherein the step of utilizing an algorithm includes storing in the database processed queries based upon a user-defined criterion.

103. The method of claim 102 wherein the user-defined criterion includes an expiration date.

104. The method of claim 60 wherein the step of utilizing an algorithm further comprises transferring the stored processed data to at least one secondary database utilizing the algorithm.

105. The method of claim 104 wherein the step of transferring the stored processed data to at least one secondary database is performed in real-time.

106. The method of claim 104 wherein the step of transferring the stored processed data to at least one secondary database is performed in batch.

107. A method for separating previously matched records, the method comprising the steps of:

determining whether a particular identifier in at least one record representing at least

one entity is one of:

common across records representing at least two different entities and generally distinctive of a record representing a particular entity; and

separating records that were previously matched based on a particular identifier if the particular identifier is determined to be common across records representing at least two different entities and not generally distinctive of a record representing a particular entity.

108. The method of claim 107 further comprising prohibiting any additional matches of records based on a particular identifier if the particular identifier is determined to be common across records representing a plurality of entities and not generally distinctive of a record representing an entity.

109. The method of claim 107 further comprising the step of re-processing the separated records.

110. The method of claim 107 wherein the steps of determining whether a particular identifier is one of common across records representing at least two different entities and generally distinctive of a record representing a particular entity and separating records that were previously matched are performed in real-time.

111. The method of claim 107 wherein the steps of determining whether a particular identifier is one of common across records representing at least two different entities and generally distinctive of a record representing a particular entity and separating records that were previously matched are performed in batch.

112. A method for processing data in a database, the method comprising the steps of:

receiving data comprising at least one record having at least one identifier, each record representing at least one of a plurality of entities;

comparing the received data with at least one record stored in a database to determine the existence of a relationship in real-time;

creating a relationship record for each record stored in a database determined to reflect a relationship with at least a portion of received data in real-time; and

storing each relationship record in the database.

113. The method of claim 112 further comprising the step of creating at least one confidence indicator for each relationship record in real time.

114. The method of claim 113 wherein at least one confidence indicator indicates the likelihood of a relationship between:

an entity represented by the particular record having a relationship with the portion of the received data, and

an entity represented by the portion of the received data.

115. The method of claim 113 wherein at least one confidence indicator indicates the likelihood that:

an entity represented by the particular record having a relationship with the portion of the received data, and

an entity represented by the portion of the received data are the same.

116. The method of claims 112 or 113 further comprising the step of analyzing the relationship records to determine whether the relationship records reflect at least one relationship not previously determined.

117. The method of claim 116 wherein the step of analyzing the relationship records includes analyzing relationship records reflecting at least one level of degrees of separation.

118. The method of claim 117 wherein the step of analyzing relationship records reflecting at least one level of degrees of separation includes analyzing relationship records meeting at least one user-defined criterion.

119. The method of claim 118 wherein the step of analyzing relationship records meeting at least one user-defined criterion includes limiting the relationship records analyzed to a maximum level of degrees of separation.

120. The method of claim 118 wherein the step of analyzing relationship records meeting at least one user-defined criterion includes limiting the relationship records analyzed to relationship records that include confidence indicators greater than a minimum amount.

121. The method of claim 116 further comprising the step of issuing an alert based upon at least one user-defined alert rule.

122. The method of claim 121 wherein the step of issuing an alert based upon at least one user-defined alert rule includes having the alert communicated via electronic communication means.

123. The method of claim 122 wherein the electronic communication means comprise an e-mail system.

124. The method of claim 122 wherein the electronic communication means comprise a

telephone.

125. The method of claim 122 wherein the electronic communication means comprise a beeper.

126. The method of claim 122 wherein the electronic communication means comprise a personal digital assistant.

127. The method of claim 121 further comprising the step of:

duplicating the relationship records on at least one secondary database;

distributing received data to the at least one secondary database for analysis based upon work load criteria; and

issuing the alert meeting the criteria of a user-defined alert rule from the at least one secondary database.

128. For a system for processing data and a computer readable medium containing program instructions for execution by a computer for performing the method comprising the steps of:

receiving data comprising at least one record having at least one identifier, each record representing at least one of a plurality of entities;

utilizing an algorithm to process the received data;

storing the processed data in a database;

receiving data queries for retrieving at least a portion of the data stored in the database; and

utilizing the algorithm to process the queries.

129. The computer readable medium of claim 128 wherein the entities are people.

130. The computer readable medium of claim 128 wherein the entities are personal property.

131. The computer readable medium of claim 130 wherein the personal property are vehicles.

132. The computer readable medium of claim 128 wherein the entities are real property.

133. The computer readable medium of claim 128 wherein the entities are organizations.

134. The computer readable medium of claim 128 wherein the entities are chemical compounds.

135. The computer readable medium of claim 128 wherein the entities are organic compounds.

136. The computer readable medium of claim 128 wherein the entities are proteins.

137. The computer readable medium of claim 128 wherein the entities are biological structures.

138. The computer readable medium of claim 128 wherein the entities are biometric values.

139. The computer readable medium of claim 128 wherein the entities are atomic structures.

140. The computer readable medium of claim 128 wherein the method further comprises the step of converting the received data into a standardized message format prior to utilizing an algorithm to process the received data.

141. The computer readable medium of claim 128 wherein the step of utilizing an algorithm to process the received data includes retaining an attribution of each record.

142. The computer readable medium of claim 141 wherein the step of retaining an attribution of each record includes retaining an identity of:

a source system providing each record and

a unique identifier representing record in the source system.

143. The computer readable medium of claim 141 wherein the step of retaining an attribution of each record includes retaining an identity of a query system and a particular user.

144. The computer readable medium of claim 128 wherein the step of utilizing an algorithm to process the received data includes analyzing the received data prior to one of storage in the database and query in the database.

145. The computer readable medium of claim 144 wherein the step of analyzing the received data prior to one of storage in the database and query in the database includes comparing at least one of the identifiers against one of:

a user-defined criterion and

at least one data set in one of the database and a list.

146. The computer readable medium of claim 145 wherein the compared identifier is a name of at least one of the plurality of entities and the data set is in a names root list.

147. The computer readable medium of claim 145 wherein the compared identifier is an address of at least one of the plurality of entities and the data set is in an address list.

148. The computer readable medium of claim 145 wherein the step of comparing at least one of the identifiers against a user-defined criterion includes formatting at least one identifier in accordance with a user-defined standard.

149. The computer readable medium of claim 144 wherein the step of analyzing the received data prior to one of storage in the database or query in a database includes enhancing the received data.

150. The computer readable medium of claim 149 wherein the step of enhancing the received data includes:

querying at least one data set in one of a database and list for additional identifiers for the received data, and

supplementing the received data with the additional identifiers.

151. The computer readable medium of claim 150 wherein the step of querying at least one data set includes:

at least one data set being in at least one database utilizing the algorithm to query additional databases to locate additional identifiers relating to at least one of the received identifiers; and

supplementing the received data with the additional identifiers located in at least one additional database.

152. The computer readable medium of claim 144 wherein the step of analyzing the received data prior to one of storage in the database and query in the database includes creating hash keys of the identifiers.

153. The computer readable medium of claim 128 wherein the step of utilizing an algorithm to process received data includes storing in the database processed queries based upon a user-defined criterion.

154. The computer readable medium of claim 153 wherein the user-defined criterion includes an expiration date.

155. The computer readable medium of claim 128 wherein the steps of receiving data comprising at least one record having at least one identifier, each record representing at least one of a plurality of entities, utilizing an algorithm to process the received data, and storing the processed data in a database are performed in real-time.

156. The computer readable medium of claim 128 wherein the steps of receiving data comprising at least one record having at least one identifier, each record representing

at least one of a plurality of entities, utilizing an algorithm to process the received data, and storing the processed data in a database are performed in batch.

157. The computer readable medium of claims 128 or 144 wherein the step of utilizing an algorithm to process the received data includes:

retrieving from the database a group of additional records having identifiers similar to the identifiers in the received data;

analyzing each identifier of the retrieved group of records for a match to at least a portion of the received data;

matching at least a portion of the received data with at least one analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities;

analyzing whether at least one identifier is included in the at least a portion of the received data that was not previously stored in the at least one analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities; and

re-analyzing each identifier of the retrieved group of records for a match to:

at least a portion of the received data and

the analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities; and

storing the matched records in the database.

158. The computer readable medium of claim 157 wherein matching at least a portion of the received data with at least one analyzed record includes assigning a persistent key.

159. The computer readable medium of claim 157 wherein the step of utilizing an algorithm to process the received data further comprises retrieving from the database an additional group of records having identifiers similar to the identifiers in:

at least a portion of the received data and

the analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities prior to re-analyzing each identifier of the retrieved group of records for a match.

160. The computer readable medium of claim 159 wherein utilizing an algorithm to

process the received data includes repeating:

retrieving from the database a group of records;

analyzing each identifier of the retrieved group of records;

matching at least a portion of the received data;

analyzing whether at least one identifier is included in the at least a portion of the received data that was not previously stored;

retrieving from the database an additional group of records; and re-analyzing each identifier of the retrieved group of records for a match until no additional matches are determined.

161. The computer readable medium of claim 157 wherein the step of utilizing an algorithm to process the received data includes:

determining whether a particular identifier is one of:

common across records representing at least two different entities and

generally distinctive of a record representing a particular entity; and

separating records that were previously matched based on a particular identifier if the particular identifier is determined after a first occurrence to be common across records representing at least two different entities and not generally distinctive of a record representing a particular entity.

162. The computer readable medium of claim 161 wherein the step of utilizing an algorithm to process the received data includes prohibiting any additional matches of records based on a particular identifier if the particular identifier is determined to be common across records representing at least two different entities and not generally distinctive of a record representing a particular entity.

163. The computer readable medium of claim 161 wherein the step of utilizing an algorithm to process the received data includes re-processing the separated records as received data.

164. The computer readable medium of claim 161 wherein the steps of determining whether a particular identifier is one of common across records representing at least two different entities and generally distinctive of a record representing a particular entity and separating records that were previously matched are performed in real-time.

165. The computer readable medium of claim 161 wherein the steps of determining whether a particular identifier is one of common across records representing at least

two different entities and generally distinctive of a record representing a particular entity and separating records that were previously matched are performed in batch.

166. The computer readable medium of claim 157 wherein the step of utilizing an algorithm to process the received data includes:

comparing the received data with at least one stored record to determine the existence of a relationship; and

creating a relationship record for each stored record determined to reflect a relationship with at least a portion of the received data.

167. The computer readable medium of claim 166 wherein the step of utilizing an algorithm to process the received data includes creating at least one confidence indicator for each relationship record.

168. The computer readable medium of claim 167 wherein the steps of comparing the received data, creating a relationship record, and creating at least one confidence indicator are performed in real-time.

169. The computer readable medium of claim 167 wherein the steps of comparing the received data, creating a relationship record, and creating at least one confidence indicator are performed in batch.

170. The computer readable medium of claim 167 wherein at least one of the confidence indicators indicates the likelihood of a relationship between:

an entity represented by the particular record having a relationship with the portion of the received data, and

an entity represented by the portion of the received data.

171. The computer readable medium of claim 167 wherein at least one of the confidence indicators indicates the likelihood that:

an entity represented by the particular record having a relationship with the portion of the received data, and

an entity represented by the portion of the received data are the same.

172. The computer readable medium of claim 167 wherein the step of utilizing an algorithm to process received data includes analyzing the relationship records to determine whether the relationship records reflect at least one relationship not previously determined.

173. The computer readable medium of claim 172 wherein the step of analyzing the

relationship records includes analyzing relationship records reflecting at least one level of degrees of separation.

174. The computer readable medium of claim 173 the step of analyzing relationship records reflecting at least one level of degrees of separation includes analyzing relationship records meeting a user-defined criterion.

175. The computer readable medium of claim 174 wherein the step of analyzing relationship records meeting a user-defined criterion includes limiting the relationship records analyzed to a maximum level of degrees of separation.

176. The computer readable medium of claim 174 wherein the step of analyzing relationship records meeting a user-defined criterion includes limiting the relationship records analyzed to relationship records that include confidence indicators greater than a minimum amount.

177. The computer readable medium of claim 172 wherein the step of utilizing an algorithm to process received data further comprises issuing an alert based upon at least one user-defined alert rule.

178. The computer readable medium of claim 177 wherein the step of issuing an alert based upon at least one user-defined alert rule includes having the alert communicated via electronic communications means.

179. The computer readable medium of claim 178 wherein the electronic communications means comprise an e-mail system.

180. The computer readable medium of claim 178 wherein the electronic communications means comprise a telephone.

181. The computer readable medium of claim 178 wherein the electronic communications means comprise a beeper.

182. The computer readable medium of claim 178 wherein the electronic communications means comprise a personal digital assistant.

183. The computer readable medium of claim 177 wherein the step of analyzing the relationship records includes:

duplicating the relationship records on at least one secondary database;

distributing received data to the at least one secondary database for analysis based upon a work load criteria; and

issuing the alert meeting the criteria of a user-defined alert rule from the at least one

secondary database.

184. The computer readable medium of claims 128 or 155 wherein the step of utilizing an algorithm to process the received data further comprises transferring the stored processed data to at least one secondary database utilizing the algorithm.

185. The computer readable medium of claim 184 wherein the step of transferring the stored processed data to at least one secondary database is performed in real-time.

186. The computer readable medium of claim 184 wherein the step of transferring the stored processed data to at least one secondary database is performed in batch.

187. For a system for processing data into and in a database and a computer readable medium containing program instructions for execution by a computer for performing the method comprising the steps of:

receiving data comprising at least one record having at least one identifier, each record representing at least one of a plurality of entities;

utilizing an algorithm to:

retrieve from a database a group of additional records having identifiers similar to the identifiers in the received data,

analyze each identifier of the retrieved group of records for a match to at least a portion of the received data,

match at least a portion of the received data with at least one analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities,

analyze whether at least one identifier is included in the at least a portion of the received data that was not previously stored in the at least one analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities; and

re-analyze each identifier of the retrieved group of records for a match to:

at least a portion of the received data and

the analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities; and

storing the matched records in the database.

188. The computer readable medium of claim 187 wherein the step of utilizing an

algorithm to the received data with at least one analyzed record includes assigning a persistent key.

189. The computer readable medium of claim 187 wherein the step of utilizing an algorithm further comprises retrieving from the database an additional group of records having identifiers similar to the identifiers in:

at least a portion of the received data and

the analyzed record of the retrieved group of records that is determined to reflect a record having identifiers representing an identical one of the plurality of entities prior to re-analyzing each identifier of the retrieved group of records for a match.

190. The computer readable medium of claims 187 or 189 wherein the step of utilizing an algorithm includes repeating:

retrieving from a database a group of additional records;

analyzing each identifier of the retrieved group of records;

matching at least a portion of the received data;

analyzing whether at least one identifier is included in the at least a portion of the received data that was not previously stored;

retrieving from the database an additional group of records; and

re-analyzing each identifier of the retrieved group of records for a match until no additional matches are determined.

191. The computer readable medium of claim 190 wherein the steps of receiving data, utilizing an algorithm and storing the matched records are performed in real-time.

192. The computer readable medium of claim 190 wherein the steps of receiving data, utilizing an algorithm and storing the matched records are performed in batch.

193. The computer readable medium of claim 187 wherein the step of utilizing an algorithm includes:

determining whether a particular identifier is one of:

common across records representing at least two different entities and

generally distinctive of a record representing a particular entity; and

separating records that were previously matched based on a particular identifier if the particular identifier is determined to be common across records representing at least two different entities and not generally distinctive of a record representing a particular entity.

194. The computer readable medium of claim 193 wherein the step of utilizing an

algorithm includes prohibiting any additional matches of records based an a particular identifier if the particular identifier is determined to be common across records representing at least two different entities and not generally distinctive of a particular entity.

195. The computer readable medium of claim 193 wherein the step of utilizing an algorithm includes re-processing the separated records as received data.

196. The computer readable medium of claim 193 wherein the steps of determining whether a particular identifier is one of common across records representing at least two entities and generally distinctive of a record representing a particular entity and separating records that were previously matched are performed in real-time.

197. The computer readable medium of claim 193 wherein the steps of determining whether a particular identifier is one of common across records representing at least two entities and generally distinctive of a record representing a particular entity and separating records that were previously matched are performed in batch.

198. The computer readable medium of claim 187 wherein the step of utilizing an algorithm includes:

comparing the received data with at least one stored record to determine the existence of a relationship; and

creating a relationship record for each stored record determined to reflect a relationship with at least a portion of the received data.

199. The computer readable medium of claim 198 wherein the step of utilizing an algorithm includes creating at least one confidence indicator for each relationship record.

200. The computer readable medium of claim 199 wherein the steps of comparing the received data, creating a relationship record, and creating at least one confidence indicator are performed in real-time.

201. The computer readable medium of claim 199 wherein the steps of comparing the received data, creating a relationship record, and creating at least one confidence indicator are performed in batch.

202. The computer readable medium of claim 199 wherein at least one of the confidence indicators indicates the likelihood of a relationship between:

an entity represented by the particular record having a relationship with the portion of

the received data, and

an entity represented by the portion of the received data.

203. The computer readable medium of claim 199 wherein at least one of the confidence indicators indicates the likelihood that:

an entity represented by the particular record having a relationship with the portion of the received data, and

an entity represented by the portion of the receive data are the same.

204. The computer readable medium of claim 199 wherein the step of utilizing an algorithm includes analyzing the relationship records to determine whether the relationship records reflect at least one relationship not previously determined.

205. The computer readable medium of claim 204 wherein the step of analyzing the relationship records includes analyzing relationship records reflecting at least one level of degree of separation.

206. The computer readable medium of claim 205 wherein the step of analyzing relationship records reflecting at least one level of degree of separation includes analyzing relationship records meeting at least one user-defined criterion.

207. The computer readable medium of claim 206 wherein the step of analyzing relationship records meeting a user-defined criterion includes limiting the relationship records analyzed to a maximum level of degrees of separation.

208. The computer readable medium of claim 206 wherein the step of analyzing relationship records meeting a user-defined criterion includes limiting the relationship records analyzed to relationship records that include confidence indicators greater than a minimum amount.

209. The computer readable medium of claim 204 wherein the step of utilizing an algorithm further comprises issuing an alert based upon at least one user-defined alert rule.

210. The computer readable medium of claim 209 wherein the step of issuing an alert based upon at least one user-defined alert rule includes having the alert communicated via electronic communications means.

211. The computer readable medium of claim 210 wherein the electronic communications means comprise an e-mail system.

212. The computer readable medium of claim 210 wherein the electronic communications

means comprise a telephone.

213. The computer readable medium of claim 210 wherein the electronic communications means comprise a beeper.

214. The computer readable medium of claim 210 wherein the electronic communications means comprise a personal digital assistant.

215. The computer readable medium of claim 209 wherein the step of analyzing the relationship records includes:

duplicating the relationship records on at least one secondary database;

distributing received data to the at least one secondary for analysis based upon a work load criteria; and

issuing the alert based upon the user-defined alert rule from the at least one secondary database.

216. The computer readable medium of claim 187 further comprising the step of converting the received data into a standardized message format prior to utilizing an algorithm.

217. The computer readable medium of claim 187 wherein the step of utilizing an algorithm includes retaining an attribution of each of the identifiers.

218. The computer readable medium of claim 217 wherein the step of retaining the attribution of each record includes retaining an identity of:

a source system providing each record and

a unique identifier representing each record in the source system.

219. The computer readable medium of claim 217 wherein the step of retaining the attribution of each record includes retaining an identity of a query system and a particular user.

220. The computer readable medium of claim 187 wherein the step of utilizing an algorithm includes analyzing the received data prior to one of storage in a database and query in the database.

221. The computer readable medium of claim 220 wherein the step of analyzing the received data prior to one of storage in the database and query in the database includes comparing at least one of the identifiers against one of:

a user-defined criterion and

at least one data set in one of a database and list.

222. The computer readable medium of claim 221 wherein the compared identifier is a name of at least one of the plurality of entities and the data set is in a names root list.

223. The computer readable medium of claim 221 wherein the compared identifier is an address of at least one of the plurality of entities and the data set is in an address list.

224. The computer readable medium of claim 221 wherein the step of comparing at least one of the identifier against a user-defined criterion includes formatting at least one identifier in accordance with the user-defined criterion.

225. The computer readable medium of claim 220 wherein the step of analyzing the received data prior to one of storage in the database and query in a database includes enhancing the received data.

226. The computer readable medium of claim 225 wherein the step of enhancing the received data includes:

querying at least one data set in one of a database and list for additional identifiers for the received data; and

supplementing the received data with the additional identifiers.

227. The computer readable medium of claim 226 wherein the step of querying at least one data set includes:

at least one data set being in at least one database utilizing the algorithm to query additional databases to locate additional identifiers relating to at least one of the received identifiers; and

supplementing the received data with the additional identifiers located in at least one additional database.

228. The computer readable medium of claim 220 wherein the step of utilizing an algorithm includes creating hash keys of the identifiers.

229. The computer readable medium of claim 187 wherein the step of utilizing an algorithm includes storing in the database processed queries based upon a user-defined criterion.

230. The computer readable medium of claim 229 wherein the user-defined criterion includes an expiration date.

231. The computer readable medium of claim 187 wherein the step of utilizing an algorithm further comprises transferring the stored processed data to at least one secondary database utilizing the algorithm.

232. The computer readable medium of claim 231 wherein the step of transferring the stored processed data to at least one secondary database is performed in real-time.

233. The computer readable medium of claim 231 wherein the step of transferring the stored processed data to at least one secondary database is performed in batch.

234. For a system for separating previously matched records, a computer readable medium containing program instructions for execution by a computer for performing the method comprising the steps of:

determining whether a particular identifier in at least one record representing at least one entity is one of:

common across records representing at least two different entities and generally distinctive of a record representing a particular entity; and

separating records that were previously matched based on a particular identifier if the particular identifier is determined to be common across records representing at least two different entities and not generally distinctive of a record representing a particular entity.

235. The computer readable medium of claim 234 further comprising prohibiting any additional matches of records based on a particular identifier if the particular identifier is determined to be common across records representing a plurality of entities and not generally distinctive of a record representing an entity.

236. The computer readable medium of claim 234 further comprising the step of re-processing the separated records.

237. The computer readable medium of claim 234 wherein the steps of determining whether a particular identifier is one of common across records representing at least two different entities and generally distinctive of a record representing a particular entity and separating records that were previously matched are performed in real-time.

238. The computer readable medium of claim 234 wherein the steps of determining whether a particular identifier is one of common across records representing at least two different entities and generally distinctive of a record representing a particular entity and separating records that were previously matched are performed in batch.

239. For a system for processing data in a database, a computer readable medium containing program instructions for execution by a computer for performing the method comprising the steps of:

receiving data comprising at least one record having at least one identifier, each record representing at least one of a plurality of entities;

comparing the received data with at least one record stored in a database to determine the existence of a relationship in real-time;

creating a relationship record for each record stored in a database determined to reflect a relationship with at least a portion of received data in real-time; and

storing each relationship record in the database.

240. The computer readable medium of claim 239 further comprising the step of creating at least one confidence indicator for each relationship record in real time.

241. The computer readable medium of claim 240 wherein at least one confidence indicator indicates the likelihood of a relationship between:

an entity represented by the particular record having a relationship with the portion of the received data, and

an entity represented by the portion of the received data.

242. The computer readable medium of claim 240 wherein at least one confidence indicator indicates the likelihood that:

an entity represented by the particular record having a relationship with the portion of the received data, and

an entity represented by the portion of the received data are the same.

243. The computer readable medium of claims 239 or 240 further comprising the step of analyzing the relationship records to determine whether the relationship records reflect at least one relationship not previously determined.

244. The computer readable medium of claim 243 wherein the step of analyzing the relationship records includes analyzing relationship records reflecting at least one level of degrees of separation.

245. The computer readable medium of claim 244 wherein the step of analyzing relationship records reflecting at least one level of degrees of separation includes analyzing relationship records meeting at least one user-defined criterion.

246. The computer readable medium of claim 245 wherein the step of analyzing relationship records meeting at least one user-defined criterion includes limiting the relationship records analyzed to a maximum level of degrees of separation.

247. The computer readable medium of claim 245 wherein the step of analyzing

relationship records meeting at least one user-defined criterion includes limiting the relationship records analyzed to relationship records that include confidence indicators greater than a minimum amount.

248. The computer readable medium of claim 243 further comprising the step of issuing an alert based upon at least one user-defined alert rule.

249. The computer readable medium of claim 248 wherein the step of issuing an alert based upon at least one user-defined alert rule includes having the alert communicated via electronic communication means.

250. The computer readable medium of claim 249 wherein the electronic communication means comprise an e-mail system.

251. The computer readable medium of claim 249 wherein the electronic communication means comprise a telephone.

252. The computer readable medium of claim 249 wherein the electronic communication means comprise a beeper.

253. The computer readable medium of claim 249 wherein the electronic communication means comprise a personal digital assistant.

254. The computer readable medium of claim 248 further comprising the step of:

duplicating the relationship records on at least one secondary database;

distributing received data to the at least one secondary database for analysis based upon work load criteria; and

issuing the alert meeting the criteria of a user-defined alert rule from the at least one secondary database.