

MICROFICHE APPENDIX

Attached as an microfiche Appendix is the MASTER.XLS spreadsheet specifying the required content of the Structured entity relationship model, the Grammar Rules, the Application Data Tables, Application Definitions, and Research rules as used in the preferred embodiment. The MASTER.XLS spreadsheet specifies the required content of the Structured entity relationship model. The Grammar Rules are those rules used in the preferred embodiment, as defined by the Grammar specification above. The Application Data Tables describe all the database tables that are used to store the actual records created in the Structured entity relationship model. The tables in the Application Data Tables—schema Report with the prefix “L_” contain details of the field lookup values such as the list of valid entries for the Title field on the Nominal entity. Tables with the prefix “T_” contain details of the entities and documents such as Nominal or Statement. Tables with the prefix “W_” are database views, rather than tables, and are used by the Analysis Tool. The Application Definitions describe all the database tables used by the Definitions Server that contain all the information that the application needs to be able to run. The records in these tables describe each and every document, entity, and form and their interrelationships held in the Application Data Tables which store the Structured entity relationship model.

What is claimed is:

1. A method for combining types of items of information from a plurality of text-based information sources in a plurality of formats, into items in a database, with
 - at least one information source including unstructured written text and structured text, and
 - at least one item of information including one or more attributes, with each attribute having at least one value,
 the method comprising the steps of:
 - for each information source,
 - extracting and organizing items of information from the structured and unstructured written text of the information source, to generate an index plan, with each item in the plan organized as a hierarchic data structure representing the item’s attributes, their values and the locations of the text in the information source supporting those values; and
 - researching and consolidating the items of information in the index plan into items in the database to avoid duplication of items in the database.
2. The method as recited in claim 1 wherein the items can be edited by a user.
3. The method as recited in claim 1 wherein each information source can include zero or more items of information.
4. The method as recited in claim 1 wherein each information source can include zero or more types of items of information.

5. The method as recited in claim 1 wherein the step of researching includes the step of automatically formulating and applying a SOL query.
6. The method as recited in claim 1 wherein the step of researching includes the step of creating an item in the database if that item in the index plan is not in the database.
7. The method as recited in claim 1 wherein the locations of the text in the information source supporting values in the items of the index plan are preserved in the database.
8. The method as recited in claim 1 wherein each item in the database can be traced to at least one item in an index plan.
9. The method as recited in claim 1 further comprising the step of validating the integrity of the information in the items.
10. A system for combining types of items of information from a plurality of text-based information sources in a plurality of formats, into items in a database, with
 - at least one information source including unstructured written text and structured text, and
 - at least one item of information including one or more attributes, with each attribute having at least one value,
 the system comprising:
 - an extractor configured to extract and organize items of information from the structured and unstructured written text of each information source, to generate an index plan, with each item in the plan organized as a hierarchic data structure representing the item’s attributes, their values and the locations of the text in the information source supporting those values; and
 - a research mechanism configured to research and consolidate the items of information in the index plan into items in the database to avoid duplication of items in the database.
11. The system as recited in claim 10 further comprising an input mechanism to allow a user to edit the items.
12. The system as recited in claim 10 wherein the research mechanism automatically formulates and applies a SOL query.
13. The system as recited in claim 10 wherein the research mechanism creates an item in the database if that item in the index plan is not in the database.
14. The system as recited in claim 10 wherein each item in the database can be traced to at least one item in an index plan.
15. The system as recited in claim 10 wherein the research mechanism determines the degree an item in the index plan matches an item in the database.
16. The system as recited in claim 10 further comprising a validator configured to validate the integrity of the information in the items.

* * * * *