The Rule Tuner: specification for a user interface to adjust the rule based system

24 June 2002

1. Introduction

The "Rule Tuner" is software that allows the administrator of the MoReTax system[1] to adjust parameters, thus influencing the operation of the system's inference rules, and – in effect – influencing the transmission of information between taxon-linked factual databases[2].

The inference rules of the system refer to
(i) the operations on direct relationships between two potential taxa,
(ii) the computation of the derived relationship between two potential taxa (nodes), which are connected through one or more paths, and
(iii) the evaluation of the "applicability"[3] of the information, which is originally linked to the start node, at the end node of paths.

The administrators of the system should be able to adjust and limit the application of these inference rules by means of a configuration framework for the processing of user queries in the system. The configuration consists of a list of parameters, which is stored in the database and which influence the following processes:

- The inclusion of edges is adjusted by assigning a weight to the authors of relationships and by the definition of a minimum weight, below which edges are excluded.
- Permissible paths are determined by the definition of a maximum length for the path, by the maximum number of the classification edges (higher and/or lower potential taxa - for all sources altogether or for every single source) included, and by means of assigning specific lengths to types of relationships (e.g. a congruence relationship could be ignored for the calculation of the total path length).
- The weighting of a path depending on the edge weightings is defined by the choice of a mathematical operation.
- In case of "simultaneous" paths (different paths with the same start and end nodes) the calculation of the resulting relationship is specified by the indication of the

[1] see http://www.bgbm.org/BioDivInf/Projects/MoreTax/ProjectIntro.htm
[2] see "The transmission of factual information" in Geoffroy, Güntsch & Berendsohn 2002, http://www.bgbm.org/biodivinf/Projects/MoreTax/formale_sprache_en.htm#Die Übertragbarkeit
[3] There are four categories in descending order for the applicability of factual information with respect to "their" Potential Taxon: (1) fully applicable, if the factual information applies to every element of the taxon, (2) partially applicable, if the factual information applies only to a subset, (3) doubtful applicable, if the factual information may apply to some elements of the taxon and (4) not applicable, if the factual information does not apply to any element.

relationship operator ("union" or "intersection"), the maximum weight of all paths and the permissible interval for the weights.

The actual output for users will not be determined by the inference module, but by not yet specified procedures. Nevertheless initialization parameters to specify differentiated outputs are set within this framework:

a) the lowest applicability category, as a cutoff point for output
b) output restrictions and specific comments on factual data stemming from a specific source depending on the user role and on the “quality” assigned to that source
c) output restrictions and specific comments on factual data stemming from a specific source depending on the user role and on access restrictions defined by that source for these data.

The system can store an unlimited number of different configurations, of which normally only one (the “active configuration”) is applied to external queries. Editing a (“provisional”) configuration is normally implemented by a process of copying an existing (active or inactive) configuration and fine-tuning the parameters by means of a trial-and-error process.

2. System architecture

The implementation of a remote editing tool for a centralized taxonomic database system can be based on either of the following two principles[4]:

- The client program consists of software specifically designed and implemented for the editing task.
- The client is realized with software (typically a World Wide Web browser), which is installed on the majority of computers anyway. With this approach, all data and the entire set of taxonomic rules are located at the server’s side and data entry forms are dynamically created.

The second option is considered preferable because the client software will run on any operating system and users will not have to install special software other than a browser. Over the past years upward compatibility of browsers has always been given, so that system updates due to new Operating Systems etc. would only affect the server software. However, care has to be taken to closely adhere to common WWW-standards on the server side.

[4] See Anton Güntsch, Jinling Li & Walter Berendsohn, "Design of the Internet Taxonomic Sector Editor", Euro+Med PlantBase, Dept. of Biodiversity and Laboratories, Botanic Garden and Botanical Museum Berlin-Dahlem, 2002:
http://www.euromed.org.uk/documents/SectorEditorDesignV1_0.pdf

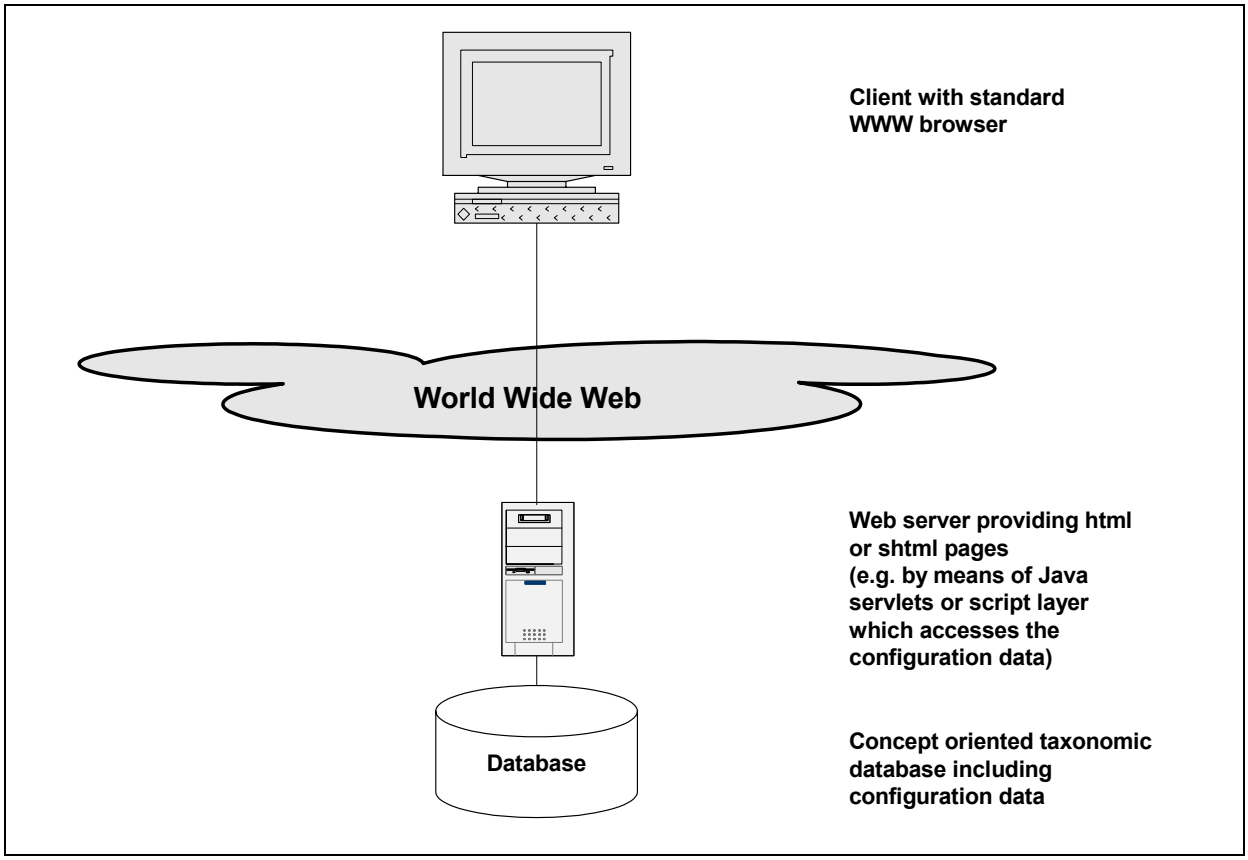

Fig.1: Rule Tuner System Architecture

For pragmatic reasons, we have decided to base further specification on a relational model for the underlying database. There are clear advantages in other data models[5], but this project aims at an implementation in the near future and – for some time to come – relational DBMS will remain the standard tool for data storage. The DBMS used must be capable of processing stored procedures, functions, and triggers so that a maximum of taxonomic data integrity can be achieved at database level.

3. Information model

The first step in the implementation of a database driven application is the definition of an appropriate information model, which has to be complex enough to meet the needs of the application and at the same time simple enough to be usable. Our taxonomic model has to incorporate nomenclatural rules and the traditional taxonomic relationships (synonymy, taxonomic hierachy, etc.). In addition, it has to be capable of representing different taxonomic views in order to enable the system to express arbitrary relationships between potential taxa. and it has to store different configurations for adjusting the inference rules.
Since several other projects are executed at the BGBM facing similar problem domains, the structure was generalized to serve (in its core) for all projects (among others, Euro+Med, IOPI / EuroCat, Med-Checklist, the Dendroflora of El Salvador and

[5] See Cédric Raguenaud, Jessie Kennedy, Peter J. Barclay, "Database Support for Taxonomy", Prometheus Technical Report #1, School of Computing, Napier University, 1999:
http://www.dcs.napier.ac.uk/~prometheus/prometheus_1/publications/pro99a.pdf

AlgaTerra). The core model presently covers nomenclatural relationships, potential taxa and their relationships, bibliographic information, and a general structure for factual data. The core model is extended in order to meet specific project requirements by means of adding further tables and relationships. For example, typification is a central subject of the AlgaTerra project, so this structure is presently being developed and tested there (it may later be added to the core). The latest versions of the core model and the MoReTax extension are documented on the Web[6].

4. The Rule Tuner interface (forms)

The user interface on the web amounts to a set of forms, which show parameter values corresponding to any configuration and allow the setting of parameter values for new (provisional) configurations, which are to be transmitted to the server and to be stored in the configuration tables of the database[7]. To provide a base for discussion, a set of example forms were developed which cover the functional needs of the system.

The main form allows the system manager to see which configuration exist and gives him the access to the central configuration managing functions. This main form could look like fig. 2.

[6] http://www.bgbm.org/BioDivInf/Projects/MoreTax/default.htm
[7] See the associated diagram at: http://www.bgbm.org/biodivinf/docs/bgbm-model/MT_Configuration.pdf

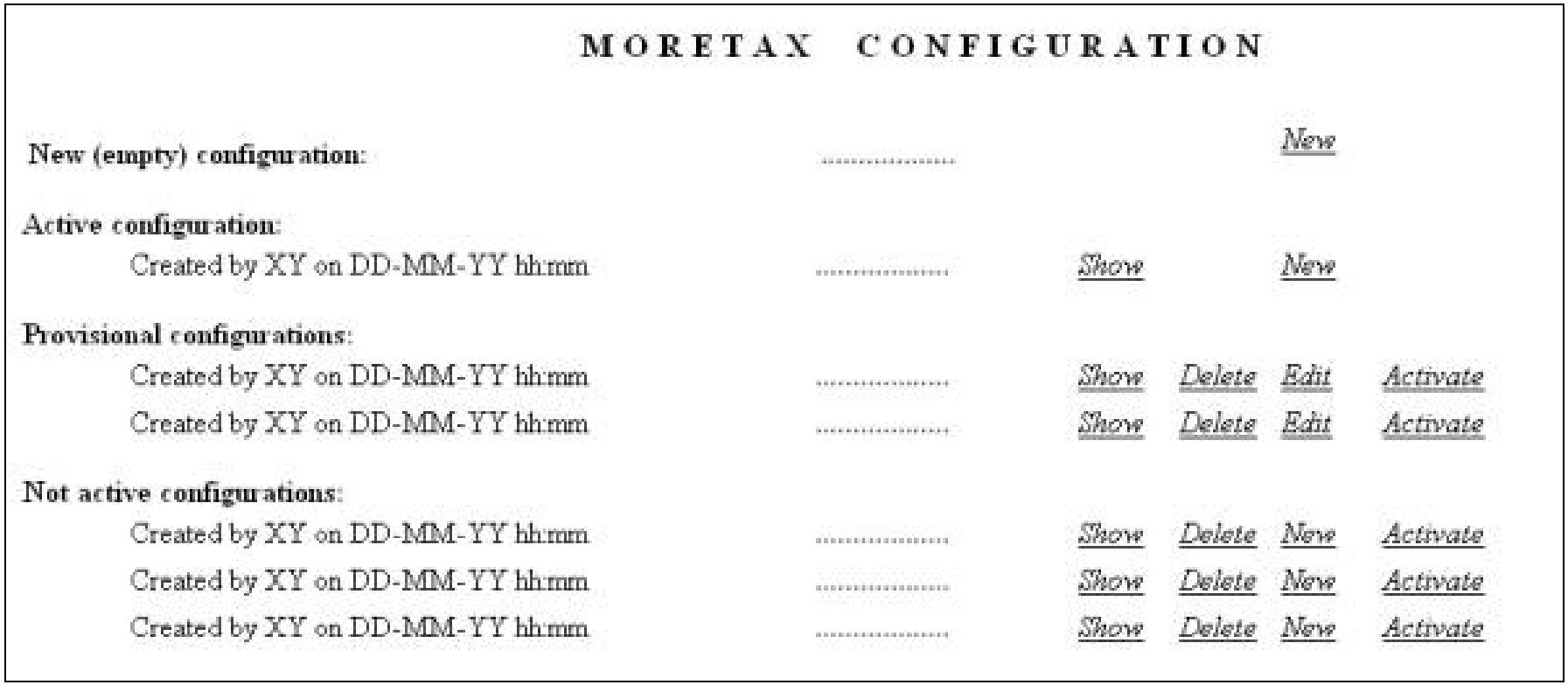

Fig. 2: Configuration Main Form; “New” creates a new provisional configuration by copying the respective existing one

Each configuration can be distinguished by means of the data referring to its creation. The system manager can choose between:

- Showing any existing configuration,
- Deleting any provisional or not active configuration (which only means that the configuration will not listed any more in this form),
- Creating a new (that means provisional) configuration either as a completely new one or based on an existing configuration,
- Editing any provisional configuration and
- Activating any provisional or not active configuration (that means substituting the current active configuration, which then becomes inactive)

4.1. The Show forms

The first form to appear after selecting the "Show" link (fig. 3) lists the parameters from the *ConfigGeneral* table of the database and gives the possibility to show further details of the configuration (data from the related tables *ConfigRelSource*, *ConfigWeight*, *ConfigFactSource* and *ConfigFact*).

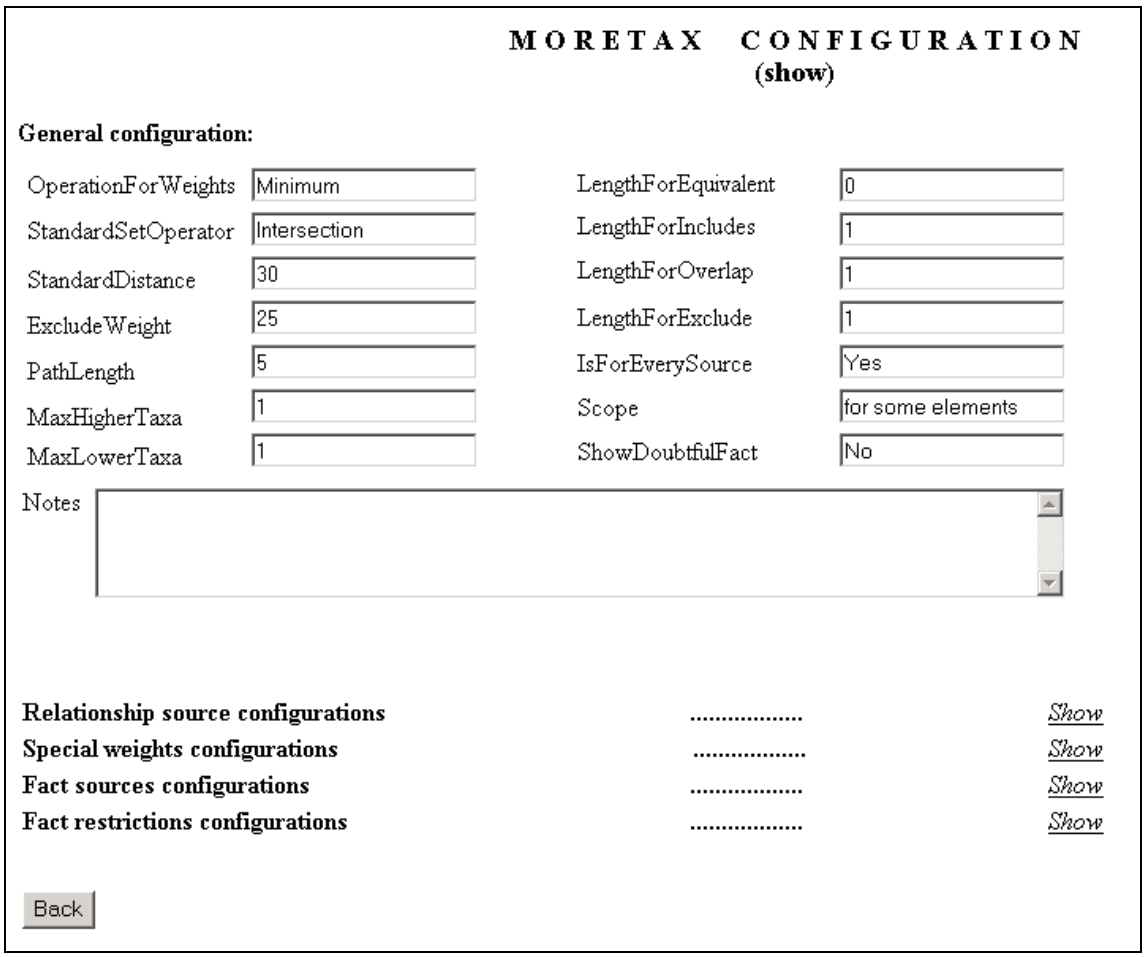

Fig. 3: Show Configuration Form (general configuration)

The form showing the relationship source configurations (fig. 4) contains the list of records of the *ConfigRelSource* table (i.e. the sources or authors who have set the relationships

between potential taxa). Each record can be distinguished through the name of the source and its creation data.

The form shown in fig. 5 contains the significant fields and their values for the chosen record from the *ConfigRelSource* table.

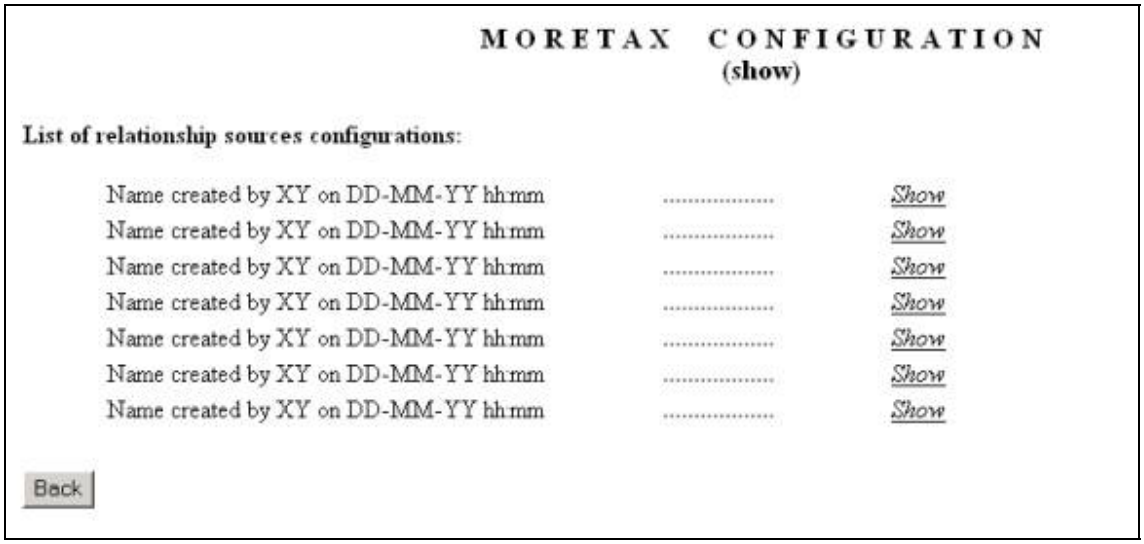

Fig. 4: List Relationship Sources Form

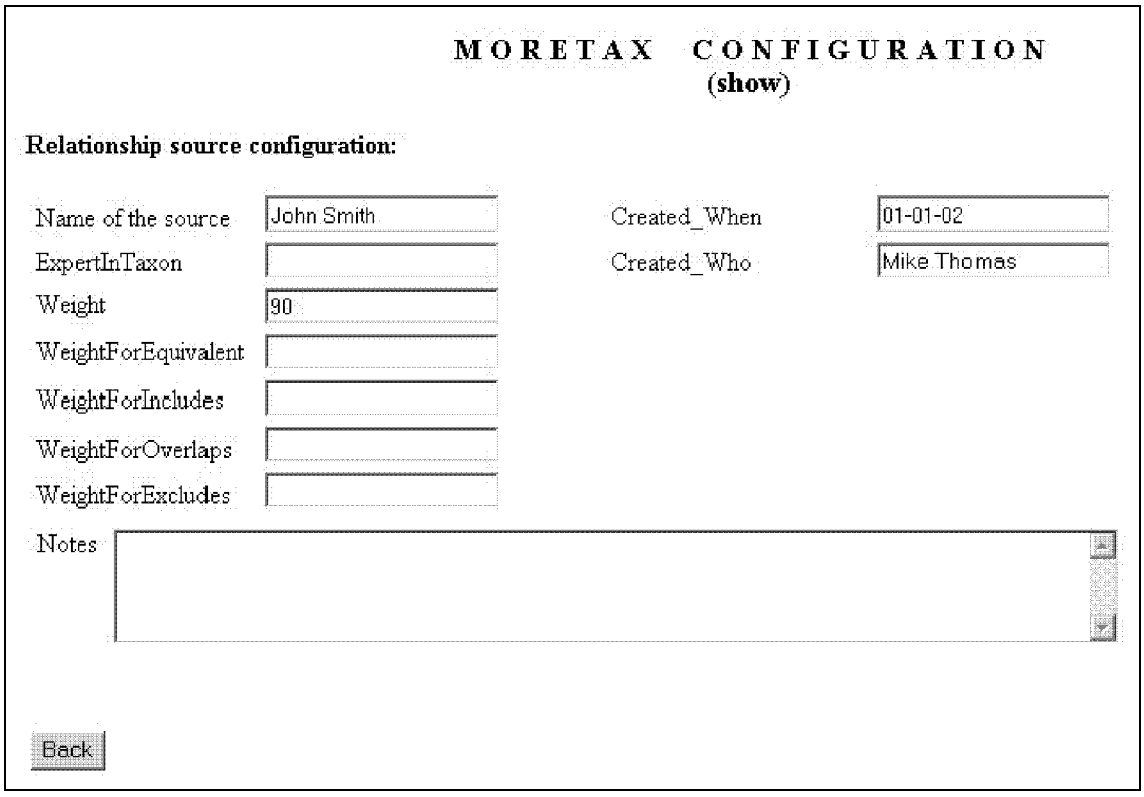

Fig. 5: Show Relationship Source Configuration Form

Back to the general configuration form (fig. 3). Choosing to show the special weights configurations results in the form shown in fig. 6, which contains on the one side the general weight treatment and on the other side the list of records of the *ConfigWeight* table (i.e. the weight treatments diverging from the general configuration). Each record can be distinguished through the weight range.

The form shown in fig. 7 contains, in addition to the general weight treatment, the significant fields and their values for the chosen record from the *ConfigWeight* table.

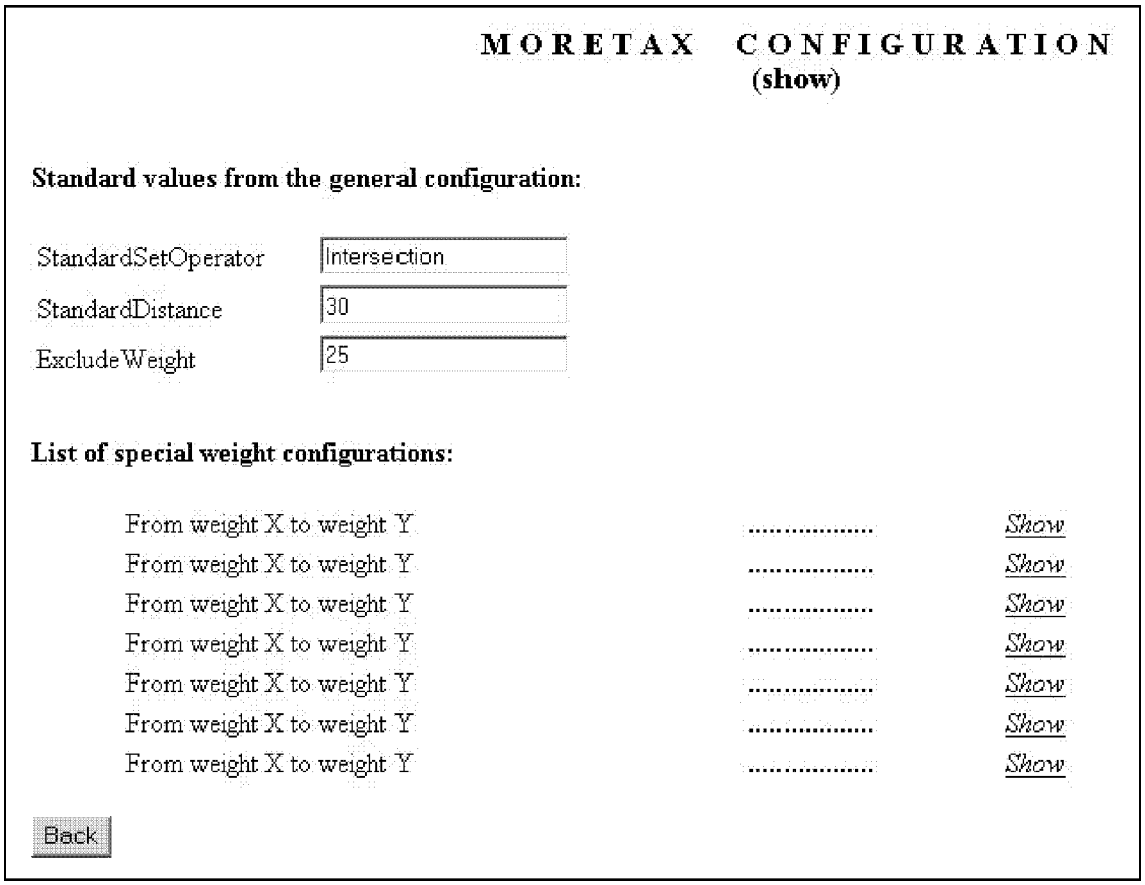

Fig. 6: List Special Weights Form

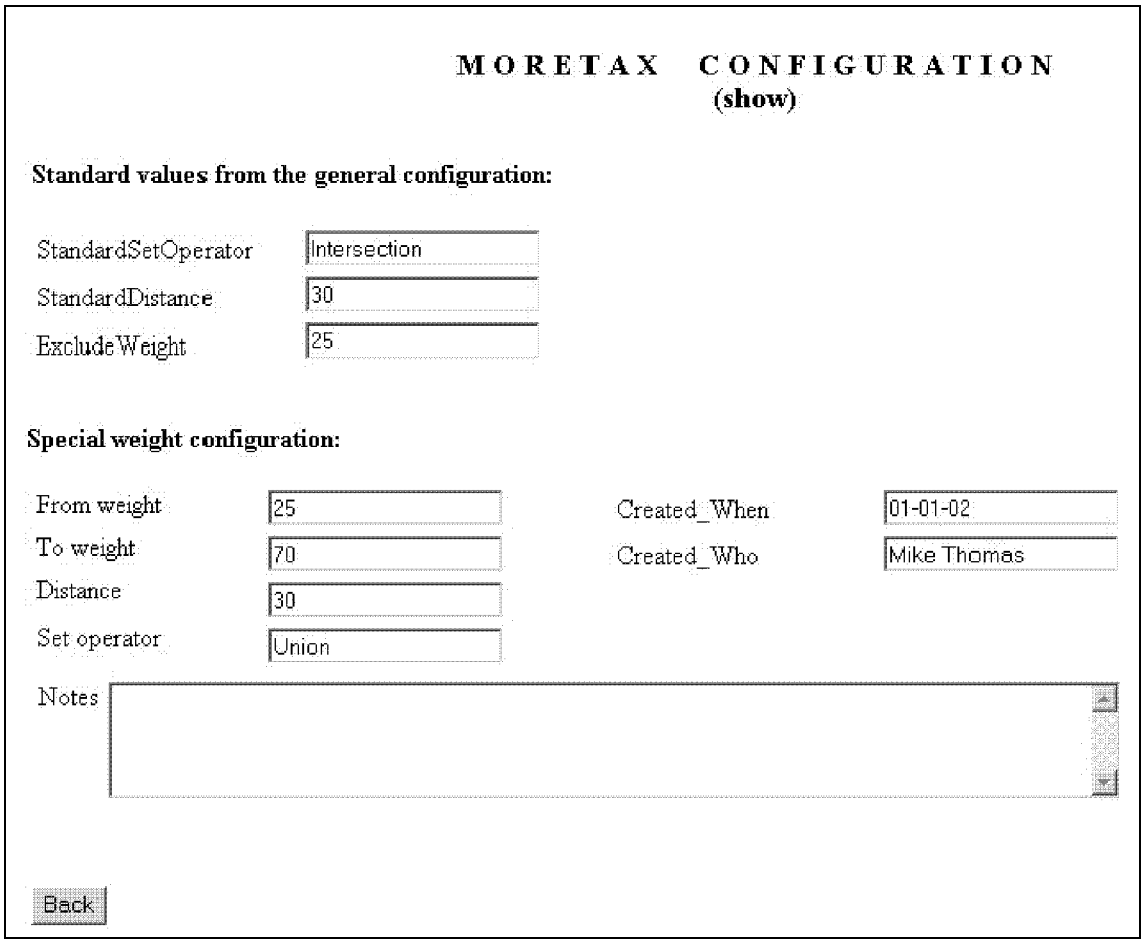

Fig. 7: Show Special Weights Configuration Form

The next option given in the general configuration (fig. 3) is to show the fact sources configurations corresponding to the selected general configuration. The resulting form (fig. 8) contains the list of records of the *ConfigFactSource* table (i.e. the output treatments for users depending on the fact source and on the user role). Each record can be distinguished through the fact source and the user role.

The form shown in fig. 9 contains, in addition to the general output treatment, the significant fields and their values for the chosen record from the *ConfigFactSource* table.

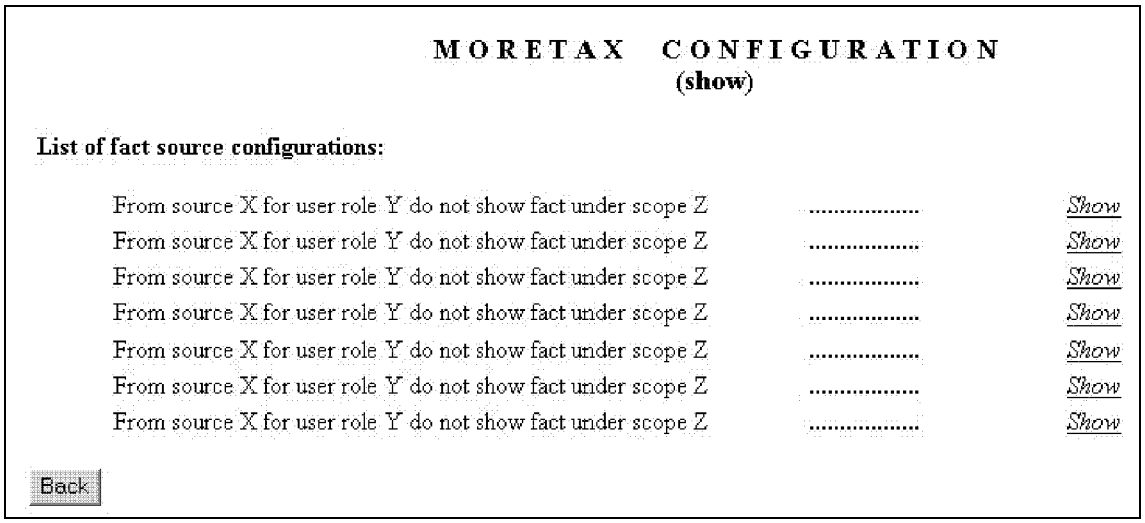

Fig. 8: List Fact Sources Form

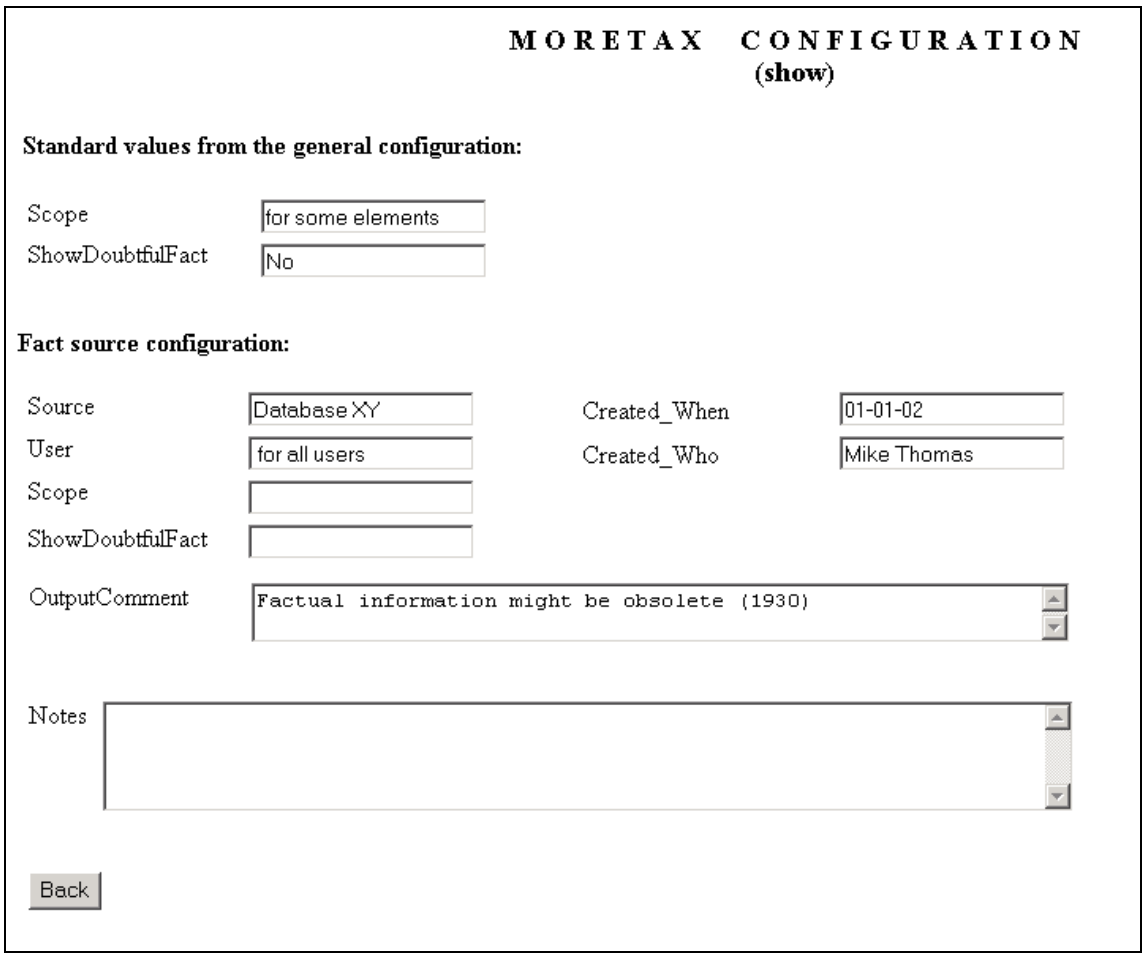

Fig. 9: Show Fact Source Configuration Form

Finally, by choosing to show the fact restrictions configurations corresponding to the selected general configuration (fig. 3), the form to appear should contain this list of records of the *ConfigFact* table (that means the output treatments for users depending on the fact

access restriction and on the user role). Each record can be distinguished through the fact restriction and the user role. Such a form is shown in fig. 10.

The form shown in fig. 11 contains, in addition to the general output treatment, the significant fields and their values for the chosen record from the *ConfigFact* table.

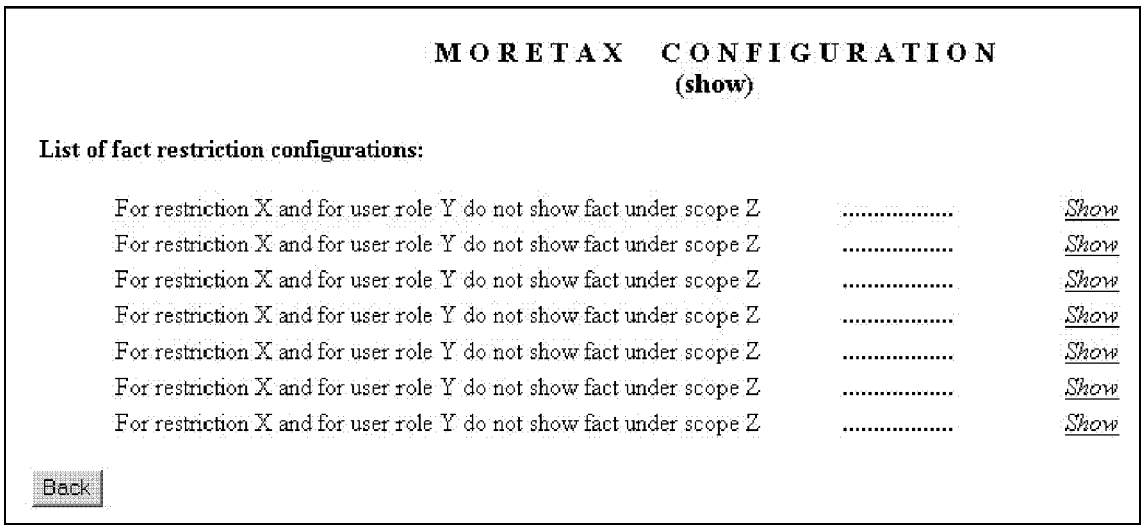

Fig. 10: List Fact Restrictions Form

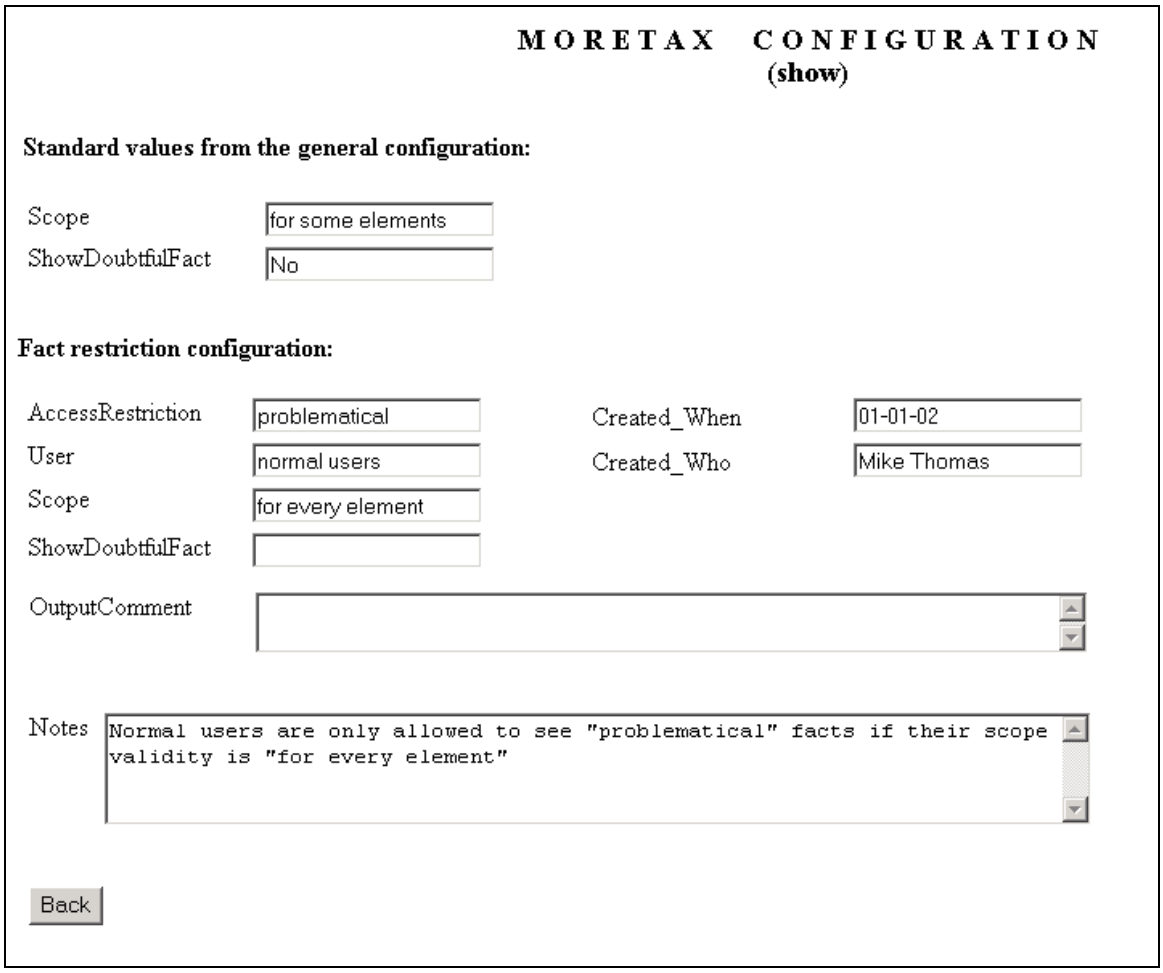

Fig. 11: Show Fact Restriction Configuration Form

4.1. The Edit forms

Edit forms are only available for provisional configurations and enable the system manager to modify the configuration either by changing values of fields within records or by adding or deleting records in the *ConfigRelSource*, *ConfigWeight*, *ConfigFactSource* or *ConfigFact* tables. The structure of these forms is very similar to the “Show” forms.

The first form to appear (fig. 12) is similar to the “Show Configuration Form” (fig. 3). It lists the parameters from the *ConfigGeneral* table of the database and gives the possibility to edit further details of the configuration (data from the related tables *ConfigRelSource*, *ConfigWeight*, *ConfigFactSource* and *ConfigFact*).

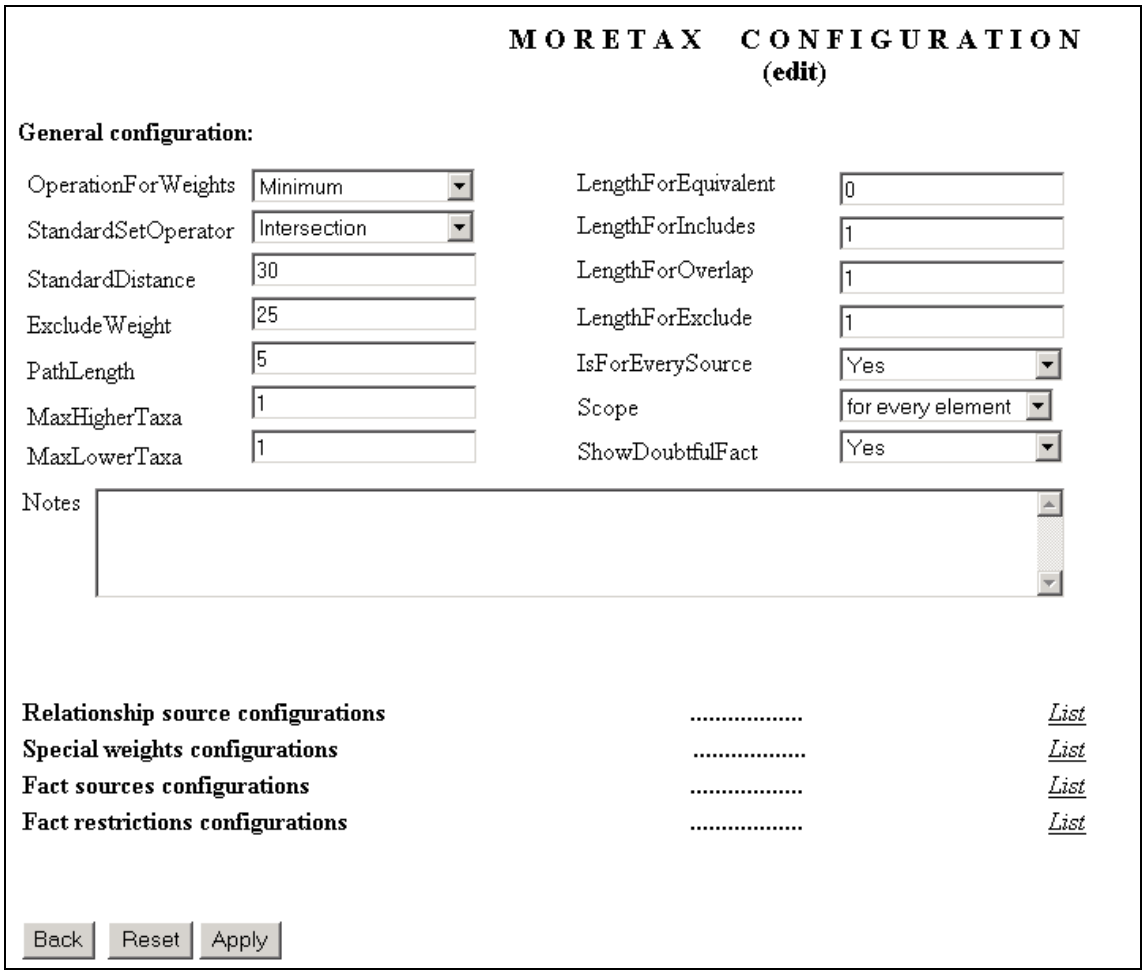

Fig. 12: Edit Configuration Form

By choosing to list the relationship source configurations corresponding to the selected general configuration, the form to appear (fig. 13) should be similar to the "List Relationship Sources" form (fig. 4). In addition, it includes the possibility

- to delete or edit any relationship source configuration already associated to the provisional configuration and
- to add new relationship source configurations to the provisional configuration either by creating a new relationship source configuration or by associating to the selected provisional general configuration a (copied) relationship source configuration from any other general configuration.

The form shown in fig. 14 contains the significant fields (and eventually their values for the chosen record from the previous form) of the *ConfigRelSource* table.

The list of relationship source configurations, which belong to other general configurations, appears in the form shown in fig. 15, from which a copy of one such record can be edited and associated with the selected provisional configuration in a new form as shown in fig. 16. The creation data cannot be edited and serve only as background information.

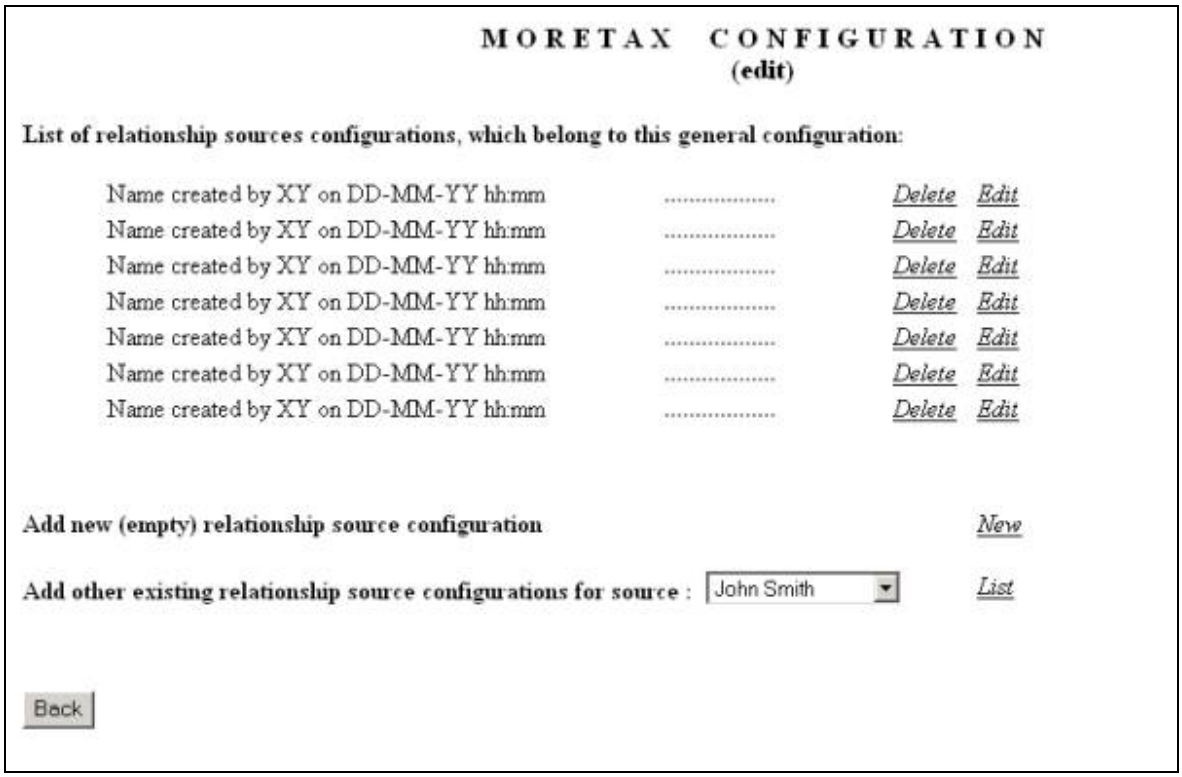

Fig. 13: Edit_list Relationship Sources Form

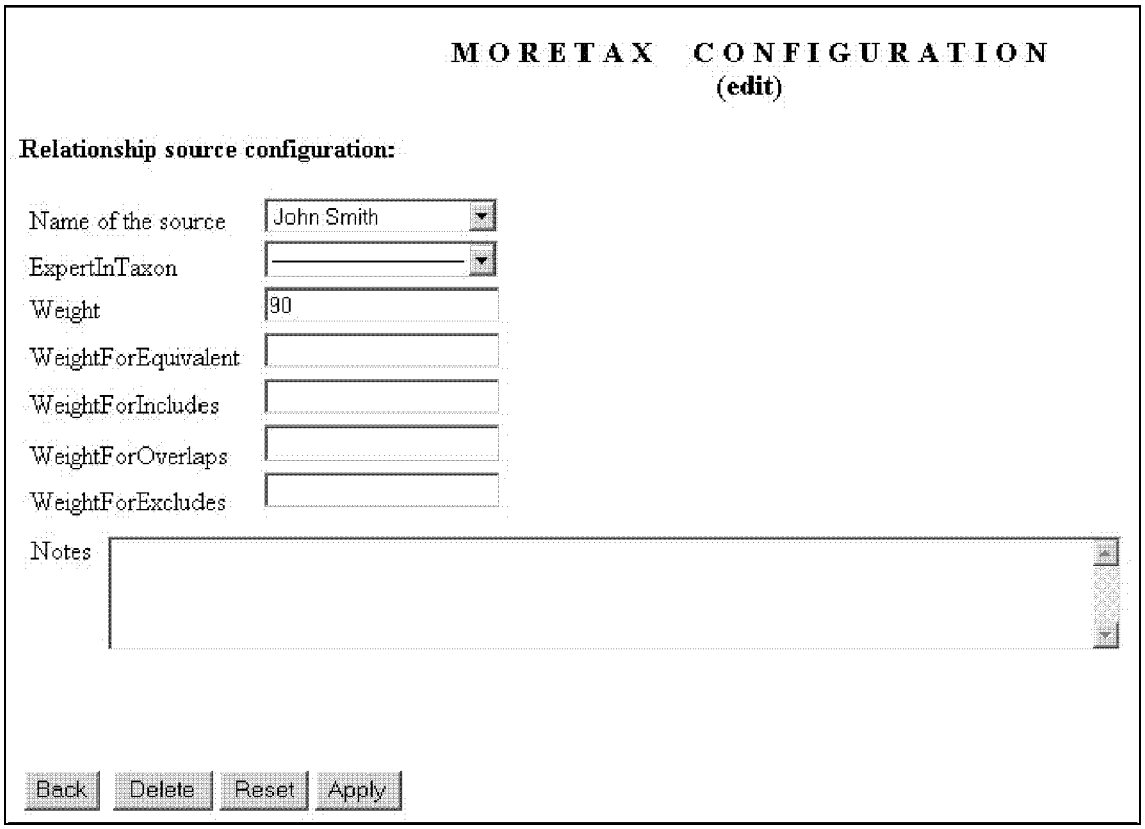

Fig. 14: Edit Relationship Source Configuration Form

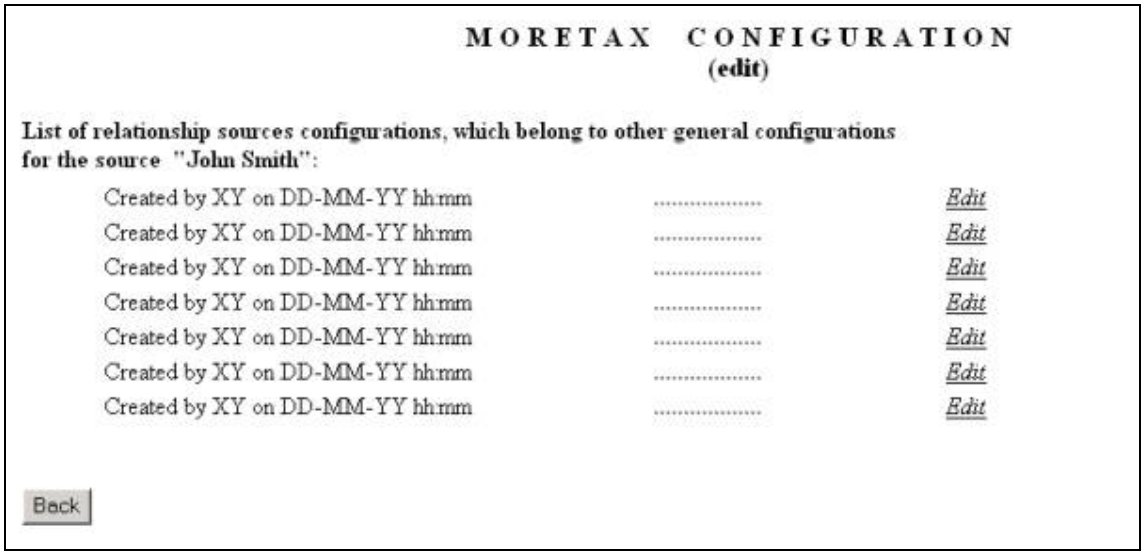

Fig. 15: Edit_list Other Existing Relationship Sources Form

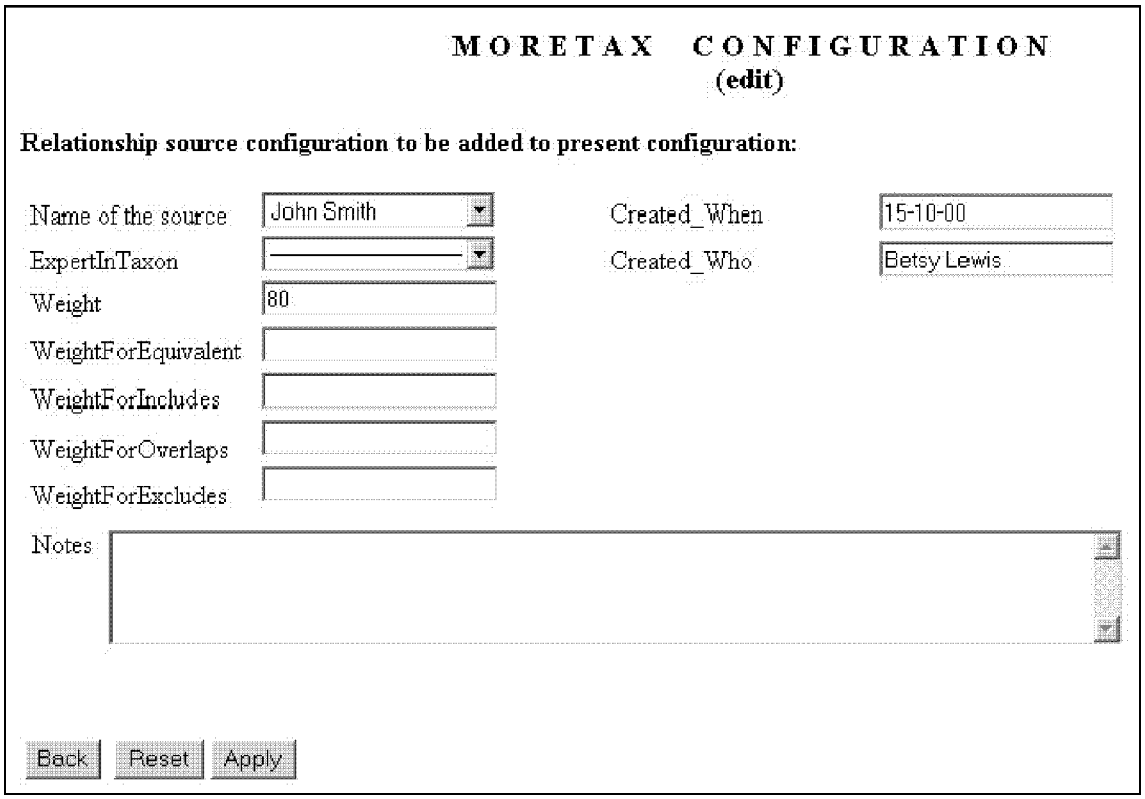

Fig. 16: Edit Other Existing Relationship Source Configuration Form

Back to the edit configuration form (fig. 12). Choosing to list the special weights configurations corresponding to the selected general configuration, the form to appear (fig. 17) should be similar to the “List Special Weights” form (fig. 6). Additionally, it includes the possibility

- to delete or edit any special weight configuration already associated to the provisional configuration and
- to add new special weight configurations to the provisional configuration either by creating a new special weight configuration or by associating to the selected provisional general configuration a (copied) special weight configuration from any other general configuration.

The form shown in fig. 18 contains the significant fields (and eventually their values for the chosen record from the previous form) for the *ConfigWeight* table.

The list of special weight configurations, which belong to other general configurations, appears in the form shown in fig. 19, from which a copy of one such record can be edited and associated with the selected provisional configuration in a new form as shown in fig. 20. The creation data cannot be edited and serve only as background information.

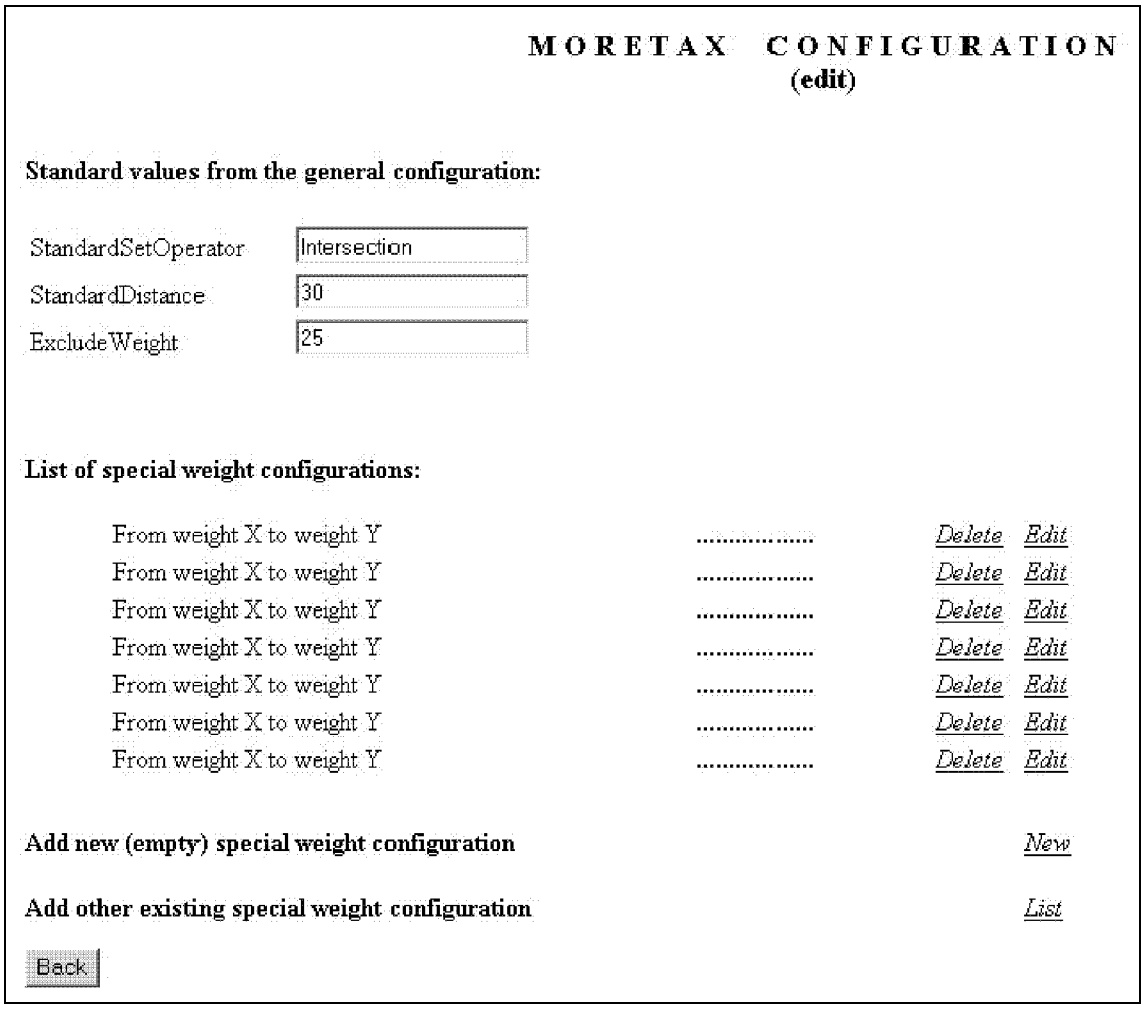

Fig. 17: Edit_list Special Weights Form

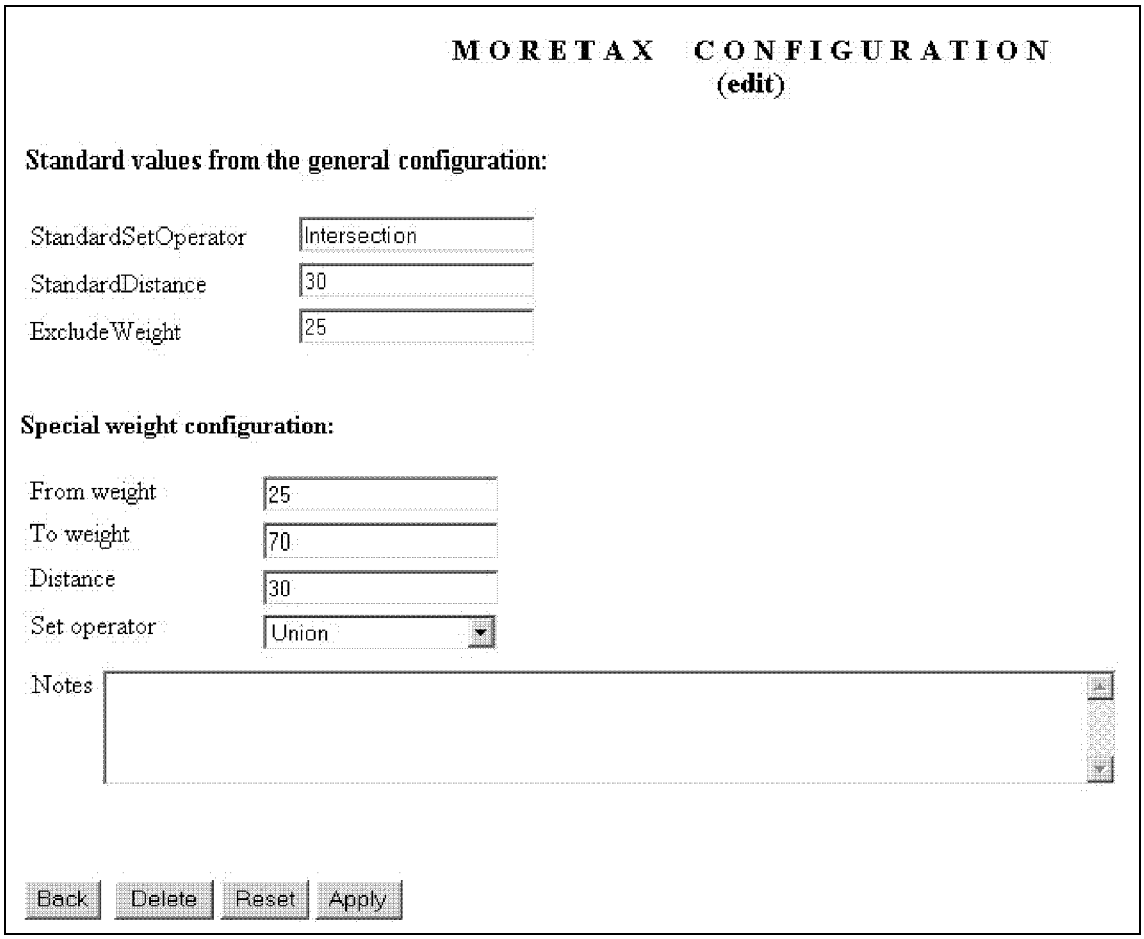

Fig. 18: Edit Special Weights Configuration Form

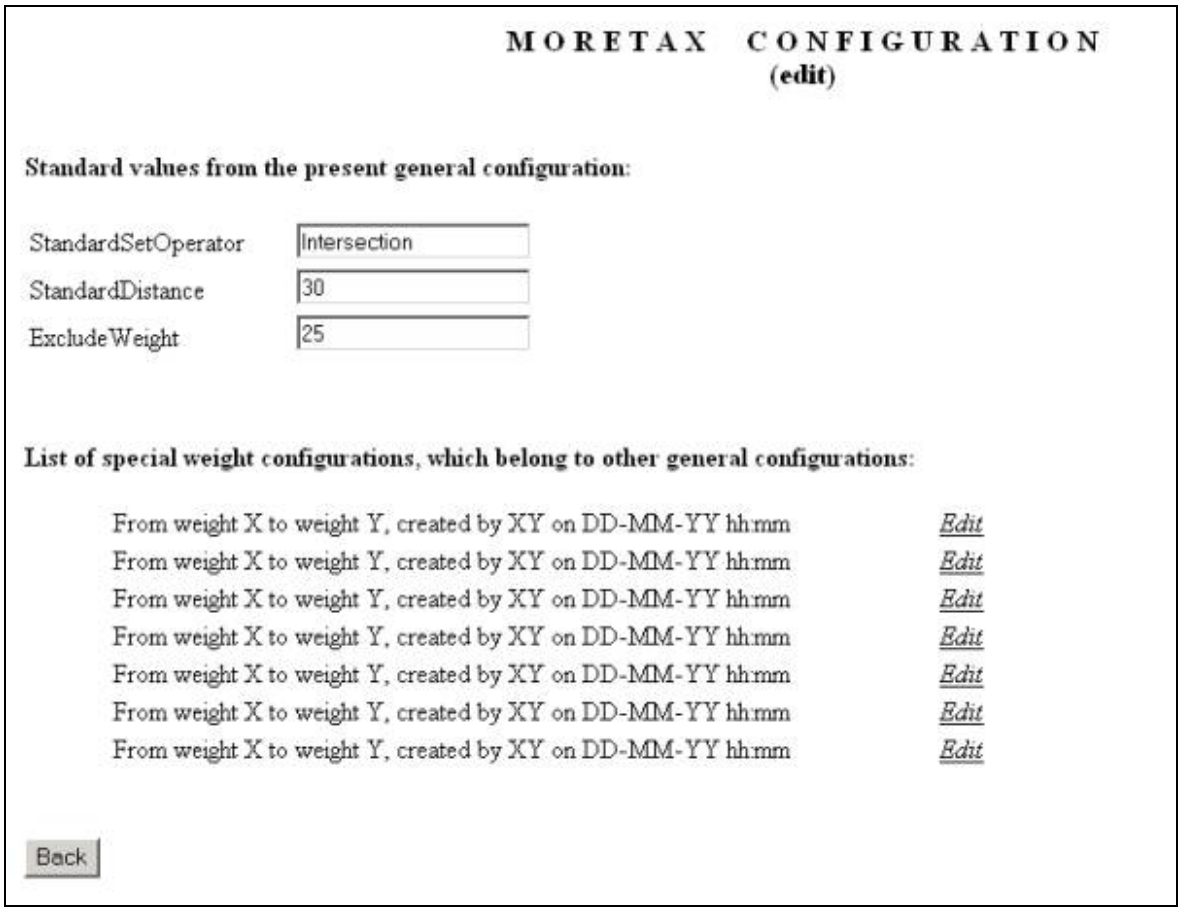

Fig. 19: Edit_list Other Existing Special Weights Form

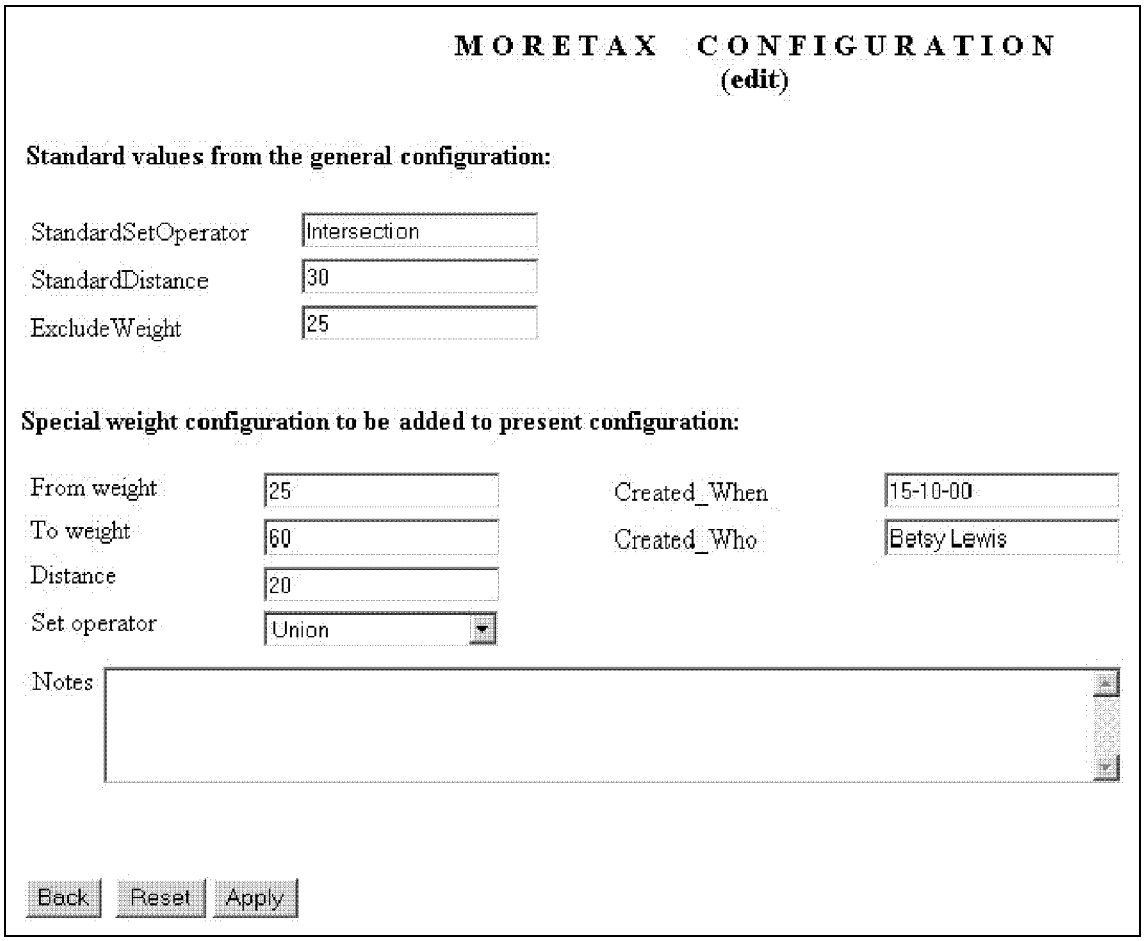

Fig. 20: Edit Other Existing Special Weights Configuration Form

Back to the edit configuration form (fig. 12). Choosing to list the fact sources configurations corresponding to the selected general configuration; the form to appear (fig. 21) should be similar to the "List Fact Sources" form (fig. 8). Additionally, it includes the possibility

- to delete or edit any fact source configuration already associated to the provisional configuration and
- to add new fact source configurations to the provisional configuration either by creating a new fact source configuration or by associating to the selected provisional general configuration a (copied) fact source configuration from any other general configuration.

The form shown in fig. 22 contains the significant fields (and eventually their values for the chosen record from the previous form) for the *ConfigFactSource* table.

The list of fact source configurations, which belong to other general configurations, appears in the form shown in fig. 23, from which a copy of one such record can be edited and associated with the selected provisional configuration in a new form as shown in fig. 24. The creation data cannot be edited and serve only as background information.

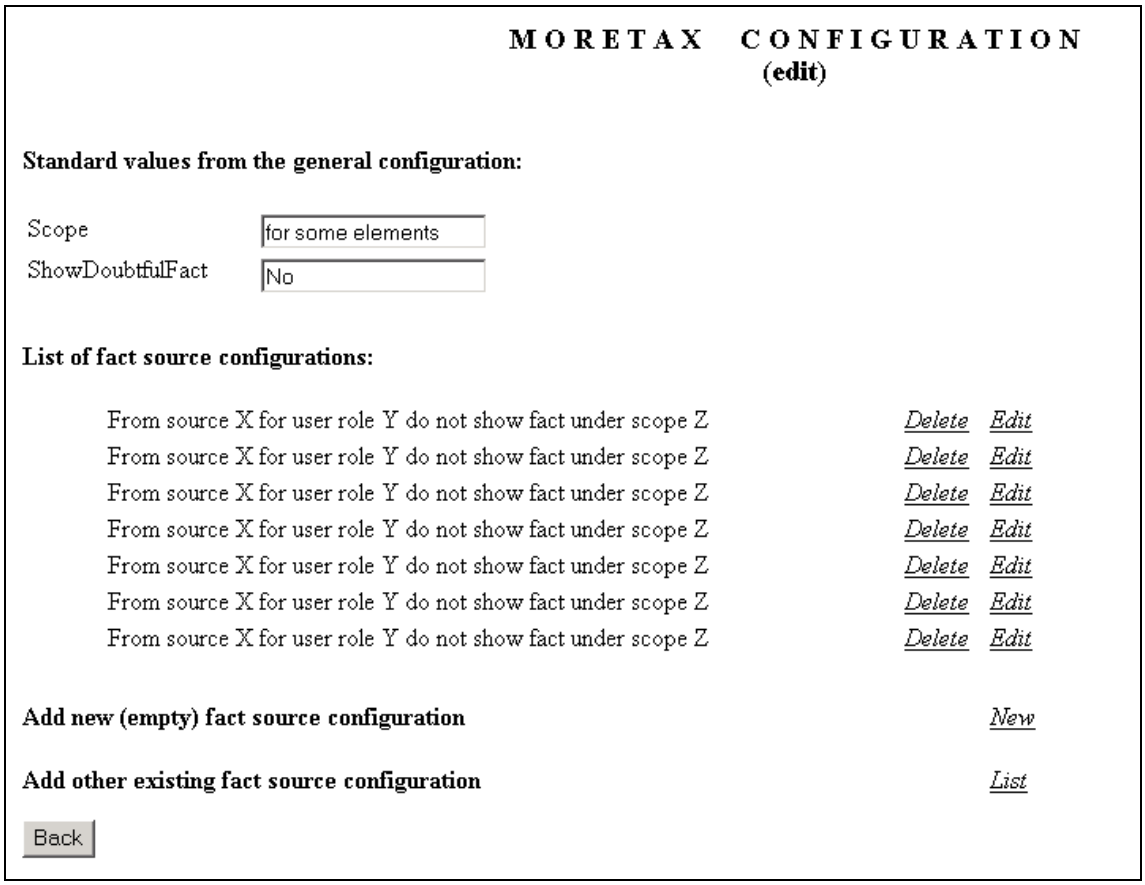

Fig. 21: Edit_list Fact Sources Form

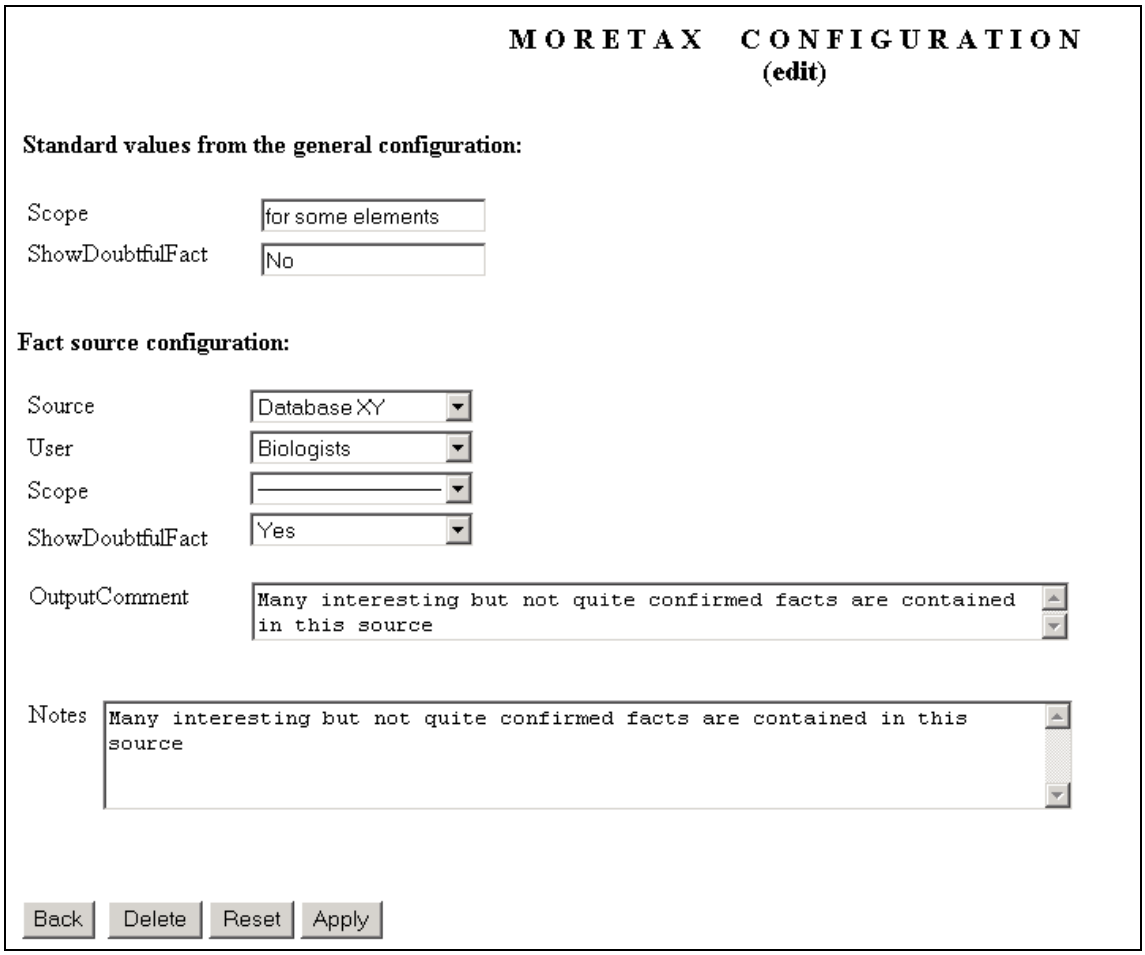

Fig. 22: Edit Fact Source Configuration Form

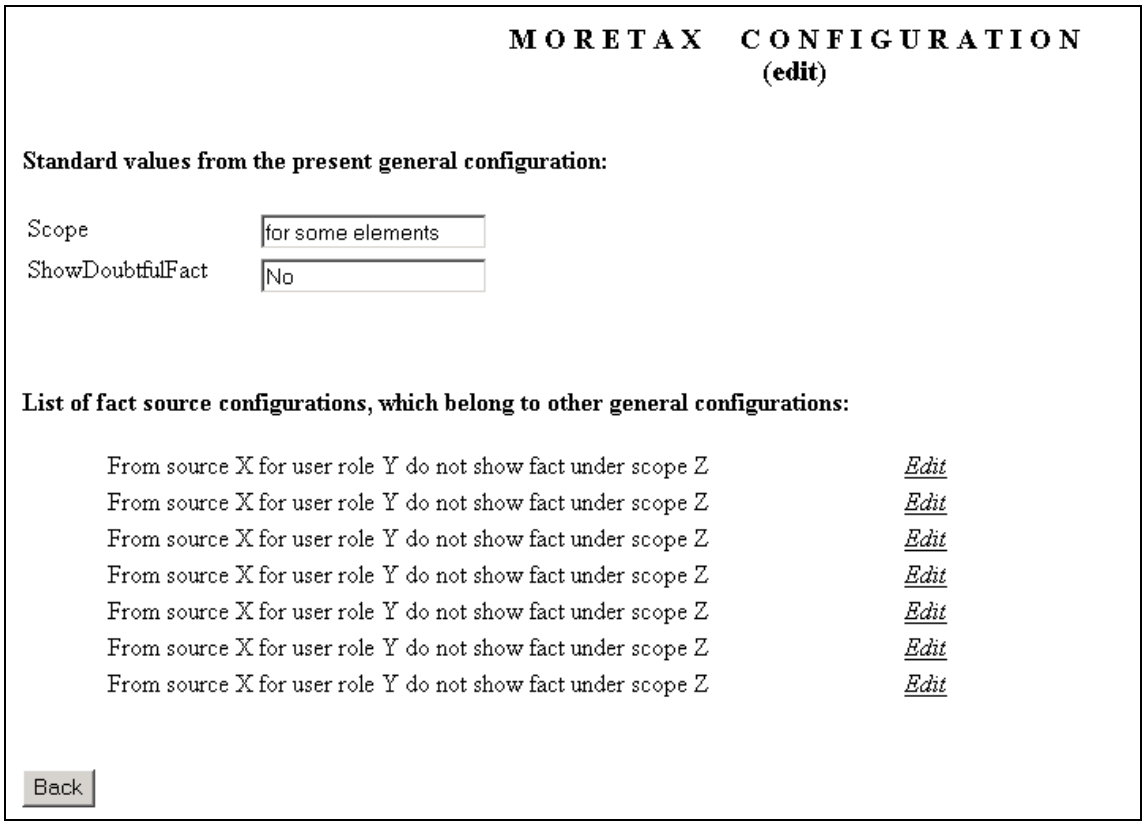

Fig. 23: Edit_list Other Existing Fact Sources Form

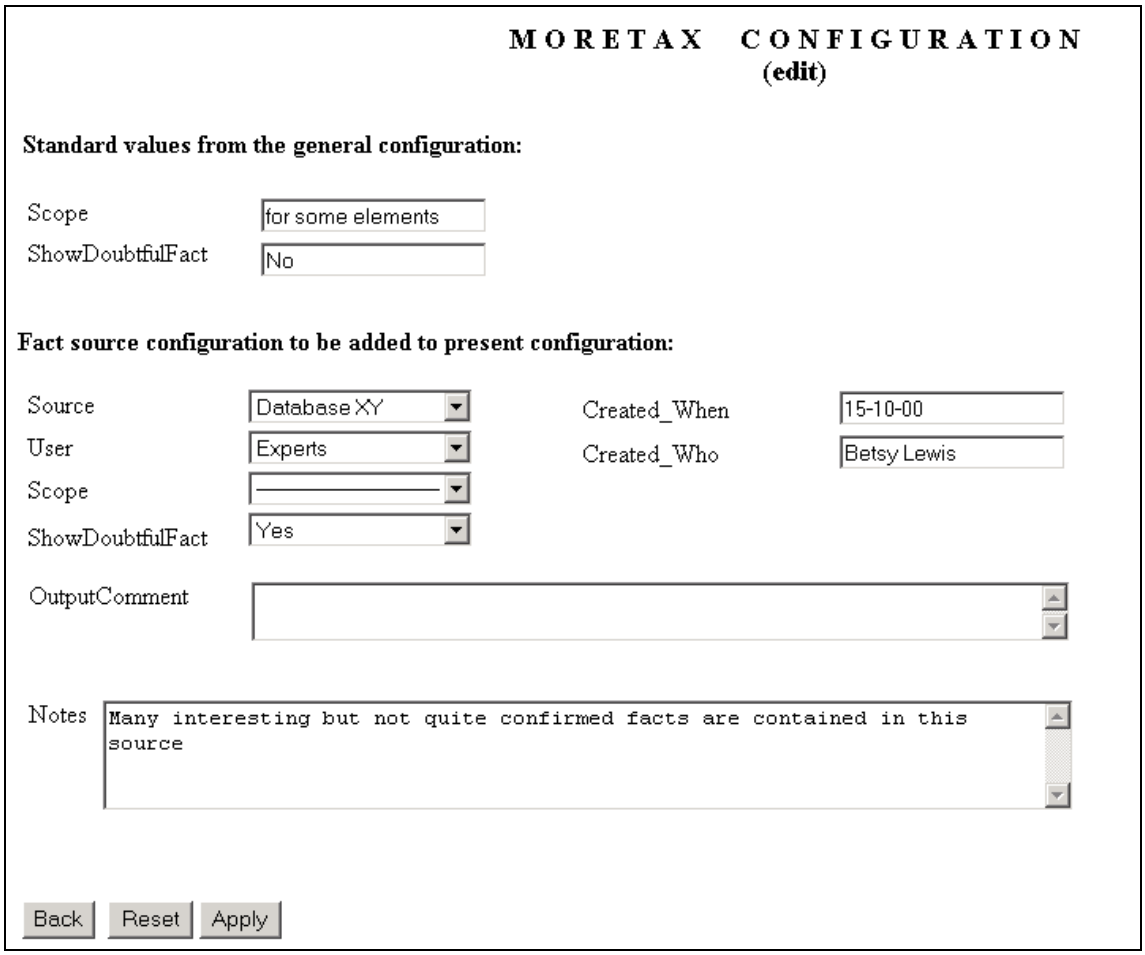

Fig. 24: Edit Other Existing Fact Source Configuration Form

Back to the edit configuration form (fig. 12). Choosing to list the fact restrictions configurations corresponding to the selected general configuration, the form to appear (fig. 25) should be similar to the "List Fact Restrictions" form (fig. 10). Additionally, it includes the possibility

- to delete or edit any fact restriction configuration already associated to the provisional configuration and
- to add new fact restriction configurations to the provisional configuration either by creating a new fact restriction configuration or by associating to the selected provisional general configuration a (copied) fact restriction configuration from any other general configuration.

The form shown in fig. 26 contains the significant fields (and eventually their values for the chosen record from the previous form) for the *ConfigFact* table.

The list of fact restriction configurations, which belong to other general configurations, appears in the form shown in fig. 27, from which a copy of one such record can be edited and associated with the selected provisional configuration in a new form as shown in fig. 28. The creation data cannot be edited and serve only as background information.

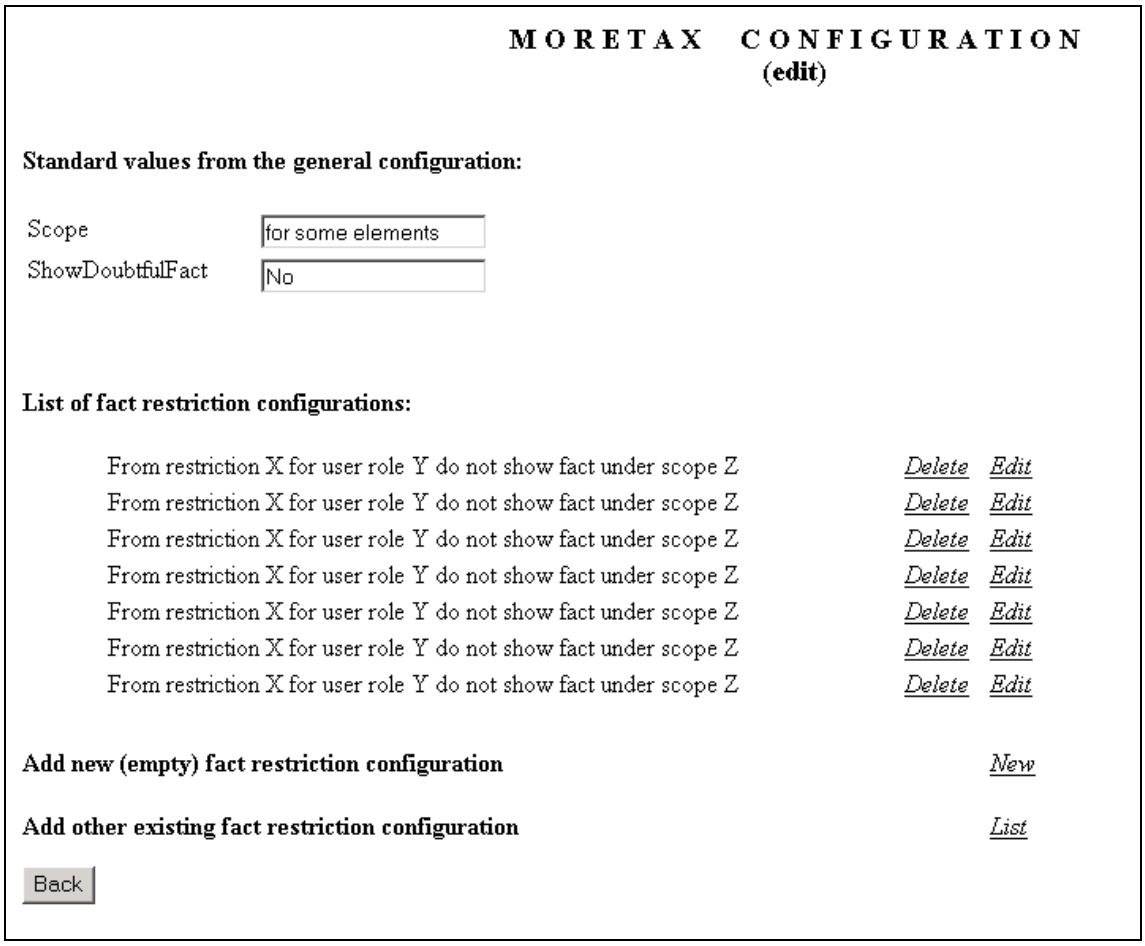

Fig. 25: Edit_list Fact Restrictions Form

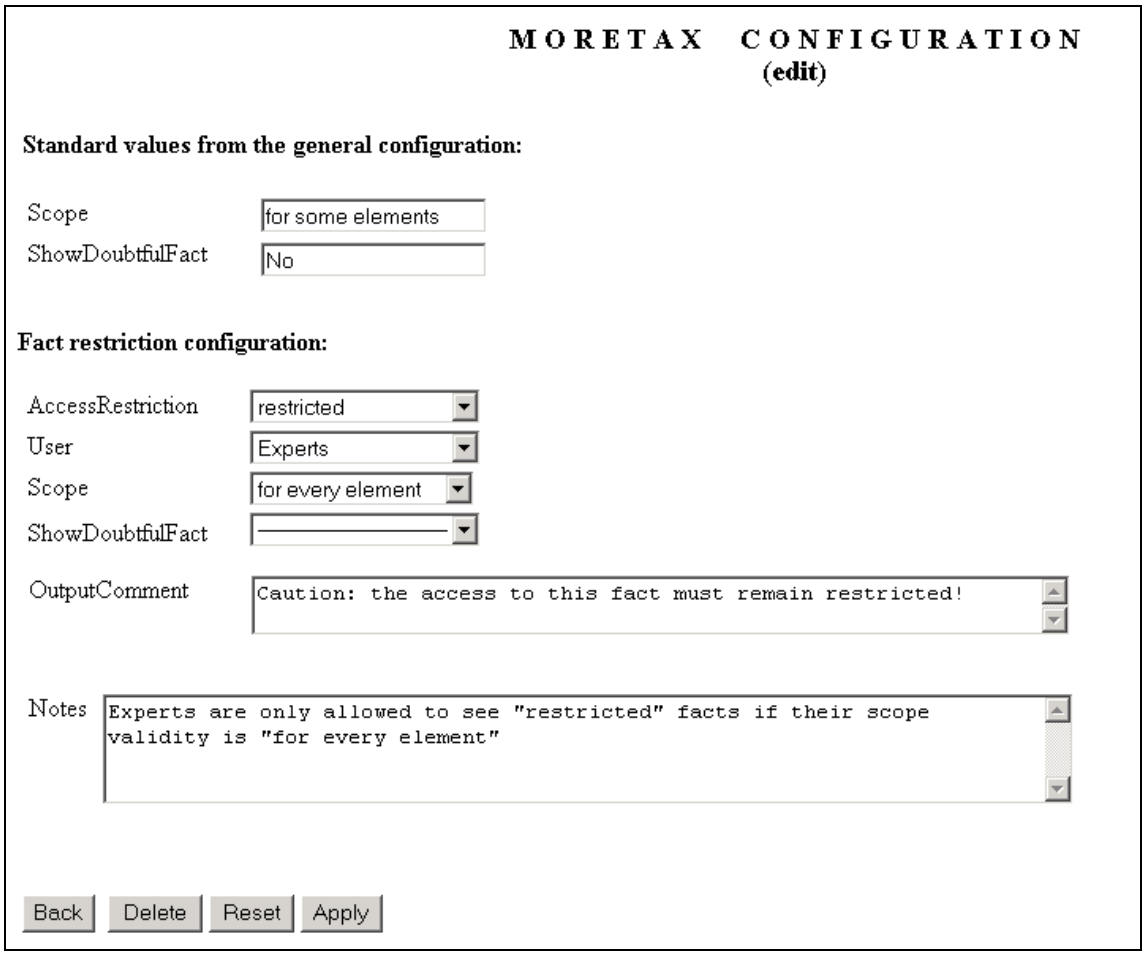

Fig. 26: Edit Fact Restriction Configuration Form

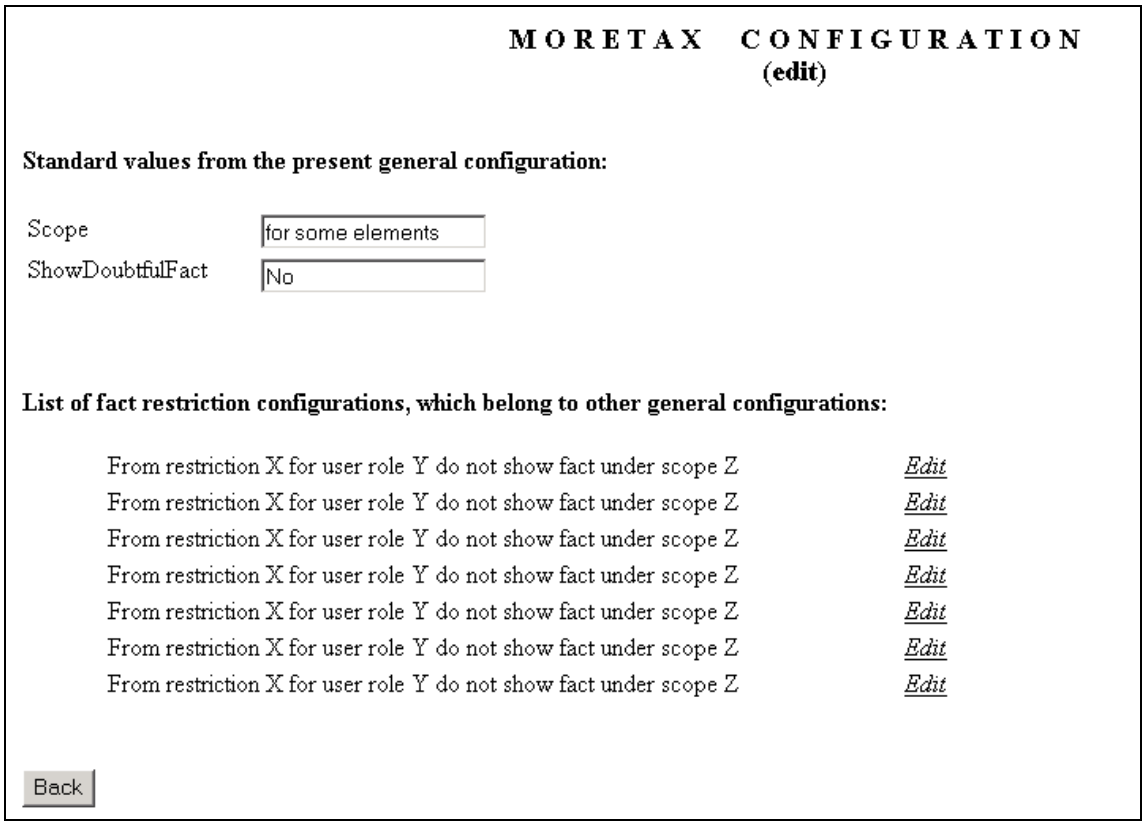

Fig. 27: Edit_list Other Existing Fact Restrictions Form

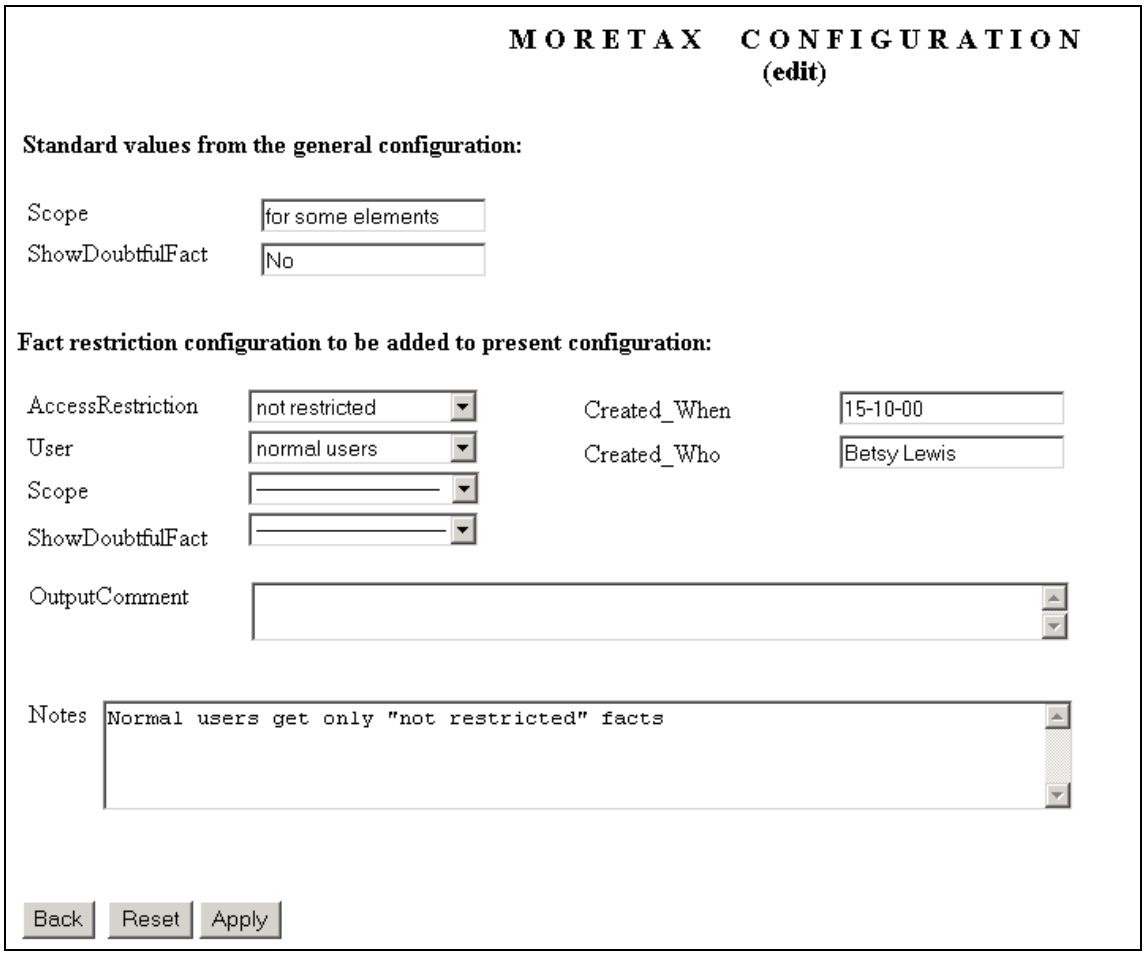

Fig. 28: Edit Other Existing Fact Source Restriction Form