

An Entity Resolution Interface for the ODINI project

An entity matching framework was designed and developed under the supervision of Dr. Tomer Sagi for publicly available oceanographic datasets.

It enables the user a simple way for reviewing datasets that from a certain collection by displaying possible relations between the datasets including an alert of possible duplicates. The system further allows resolving the relations and assists in this by providing the relevant information and a possibility to delete duplicates of a dataset from the collection in cases of records that referred to the same existing dataset or to another record.

The framework implements Qais Abou Housien's algorithm that enables to compare between records of datasets by using common attributes and referring to them using general definitions of space and time combining references of a location by coordinates and an instant or over a time interval.

It further enables comparison between different datasets in a certain collection and returns suggested relations (e.g., same-as, contained-in, extends) based on the aforementioned definitions.

In this project I was responsible for the design and implementation of the framework's user interface.

The framework consists of 5 screens relating to the entity matching framework.

The initial screen enables one to execute the algorithm on an existing collection from a list in order to find possible relations of datasets and then allows one to move to a screen that displays relations awaiting review or to another screen of relations that have been reviewed by a user.

These screens allows the user to see the suggested relation of relations awaiting review or relations that have already been reviewed. In each screen a pair of datasets from the collection are presented on a map together with their auxiliary information side by side that can help the user to resolve or edit the relation's type. Following the user decision, a possibility of deleting one dataset from the collection in case of the relation's type is resolved as similar datasets (one dataset will be deleted) or fully-overlapping records (the smaller dataset will be deleted). Additional screens allows the user to obtain an overview of the review process. One screen displays a table that allows the user to see all the relations including that have reviewed of collection's datasets according to its id and a possibility to see the relation's type, status (review/ not review) and to move to a screen displaying that relation in order to see it with auxiliary information.

In addition, there is a screen displaying a table that allows the user to see the collection's datasets, their relations and the relevant information.