Object Matching (Entity resolution) is a critical data integration task and aims at identifying semantically corresponding objects (records, instances) in one or several data sources. A typical example is the redundant and heterogeneous representation of customers in different enterprise databases. Finding corresponding customer representations is a key task, e.g., for customer relationship management or master data management, in general. On the web, finding matching objects is typically even more challenging due to the higher degrees of heterogeneity (less structured data with many text attributes, more sources, more data quality problems etc.).

We are developing comprehensive prototypes for object matching since 2006. A key idea is to support the combination of several match techniques (matchers) to improve the overall effectiveness in terms of precision and recall. The first prototype MOMA supports the construction of flexible workflows for object matching and the reuse of previous match results which are represented as instance mappings. Furthermore, MOMA not only uses the similarity of attribute values but also incorporates a powerful context matcher called neighborhood matcher.

The more recent frameworks STEM and FEVER support blocking and matching as well as the use of machine learning techniques. The machine learning approaches utilize a limited amount of training data (manually labeled correspondences) to semi-automatically find effective combinations of matchers. FEVER also supports the comparative evaluation of different match approaches for a given match task.
Object Matching / Entity Resolution

FEVER: Snapshots GUI