Persistent Quality–driven Adaptive Information Integration

Riccardo Porrini[†] [§]

Supervised by Matteo Palmonari[†]

[†]ITIS Lab, University of Milan-Bicocca, Milano (Italy) - [§]7Pixel s.r.l, Pavia (Italy)

`riccardo.porrini@disco.unimib.it`

1. Pay–as–you–go Information Integration

Pay-As-You-Go Information Integration Systems (PIIS) [HFM06] gather and integrate information coming from different sources by constantly improving and refining the integration over time. PIIS are characterized by [JFH08]:

- ★ Large amount of heterogeneous sources, changing over time (at schema and instances level);
- ★ A set of mappings between sources and target (at schema and instances level) defined by domain experts or found by (semi) automatic algorithms;
- ★ End-users who search for and browse to integrated information and whose information needs change over time.

Adapting both schema and instance mapping to end-users' dynamic information needs is crucial to support them in effectively access the integrated information. *Explicit user feedback* (i.e., from domain experts [BPE+13]) and *implicit user feedback* (i.e., inferred from end-user' observable behavior [COY12]) represent a valuable aid to mapping adaptation and can be used as an indirect measure of the overall *quality* of PIIS at a given time.

2. Joint Mapping and Feedback Management

Hypothesis

Quality defined in terms of *fitness* to end-users' information needs and *conformance* to domain experts knowledge can be used to improve integration over time

Questions

1. Can user behavior be used to assess the quality of a PIIS at a given time?
2. Can mappings maintenance and evolution be driven by continous quality evaluation?

Main Goal

A mapping and feedback management system that supports information integration maintainance and evolution

From		To
Single-user feedback loop	$\longrightarrow$	Multiuser feedback loop
One-shot self-configuring/optimized PIISs	$\longrightarrow$	Persistent and adaptive PIIS

References

[BPE+13] Khalid Belhajjame, Norman W. Paton, Suzanne M. Embury, Alvaro A. A. Fernandes, and Cornelia Hedeler. Incrementally improving dataspaces based on user feedback. *Inf. Syst.*, 38(5), 2013.

[COY12] Longbing Cao, Yuming Ou, and Philip S. Yu. Coupled behavior analysis with applications. *IEEE Trans. Knowl. Data Eng.*, 24(8), 2012.

[HFM06] Alon Y. Halevy, Michael J. Franklin, and David Maier. Principles of dataspace systems. In *PODS*, 2006.

[JFH08] Shawn R. Jeffery, Michael J. Franklin, and Alon Y. Halevy. Pay-as-you-go user feedback for dataspace systems. In *SIGMOD Conference*, 2008.

3. The Proposed Approach
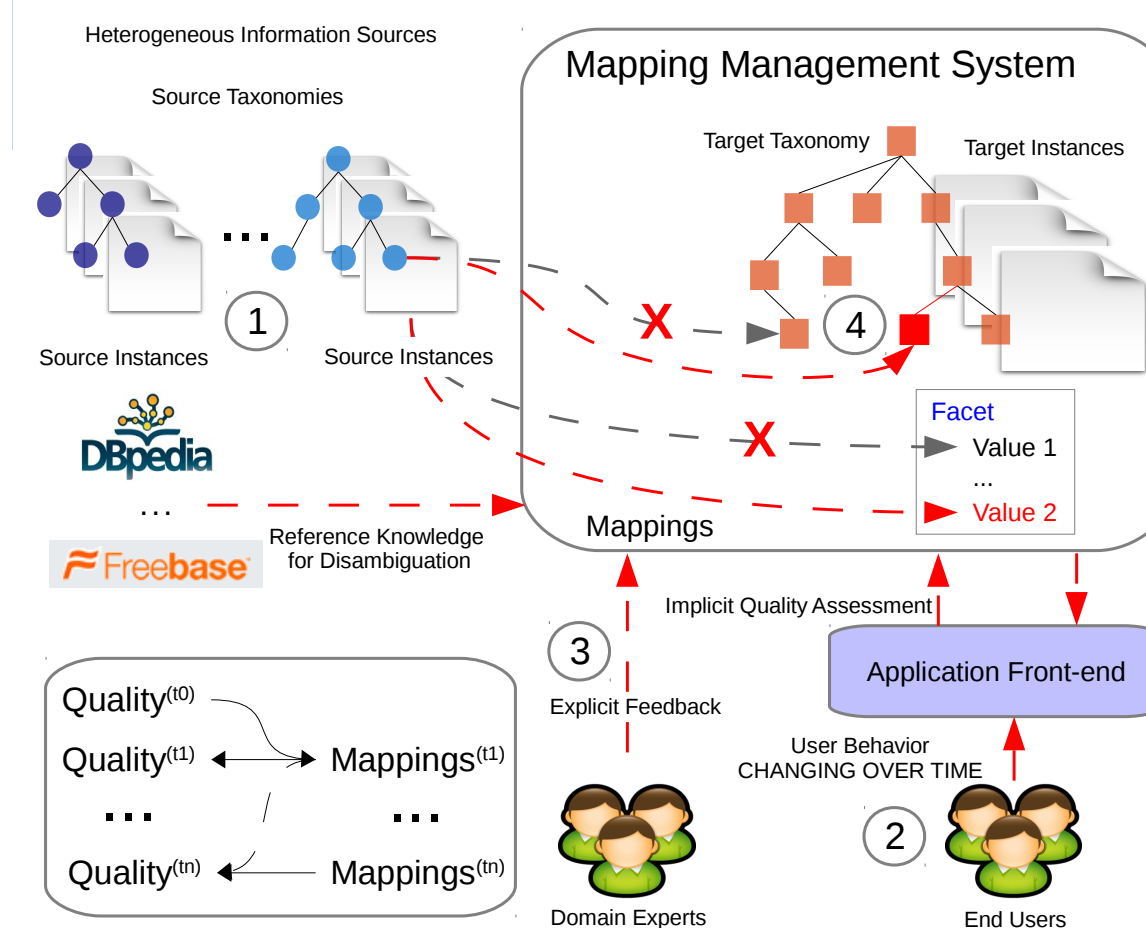
1. Weakly structured data is integrated with the help of (semi) automatic algorithms into a target source, both at schema and instance level;
2. End-users search for and browse to target instances using a target classification scheme and facets, thus providing *implicit feedback* about the fitness for use of the PIIS at a given time;
3. Domain experts provide *explicit feedback* about mappings during during their PIIS management activities;
4. *Quality* of the PIIS at a given time is assessed by considering *explicit* and *implicit feedback* and drives target schema evolution and mapping adaptation.

Relevance to Semantic Web: Ontology Matching and Link Discovery techniques can be used at many stages of the mapping management process.

4. Case Studies from Price Comparison Engines

Adaptive Mappings and Facet Management

Adaptive Mappings and Taxonomy Refinement