Advanced Information Systems project status report

Gasparella Luca

A.y. 2008/2009

Problem definition

A soaring number of services bases their functionality on a distribuited context where each local implementation is often realized separately from the others and the model used to represent data may be significatively different. In particular the problem analyzed take into the syntactic heterogeneity that fold several aspects about the definitions for representing data. The main aim of the project is to allow the execution of queries on datasources which have been described differently through two distinct schemas.

Status

References

[1] Xin Dong, Alon Halevy, *"Indexing Dataspaces"*, SIGMOD '07: Proceedings of the 2007 ACM SIGMOD international conference on Management of data (2007), pp. 43-54.

[2] AK Elmagarmid, PG Ipeirotis, VS Verykios, *"Duplicate Record Detection: A Survey"*, Knowledge and Data Engineering, IEEE Transactions on, Vol. 19, No. 1. (2007), pp. 1-16.

[3] AnHai Doan, Pedro Domingos, Alon Halevy, *"Learning to Match the Schemas of Data Sources: A Multistrategy Approach"*, Kluwer Academic Publishers, Volume 50 , Issue 3 (2003), pp. 279 - 301.