



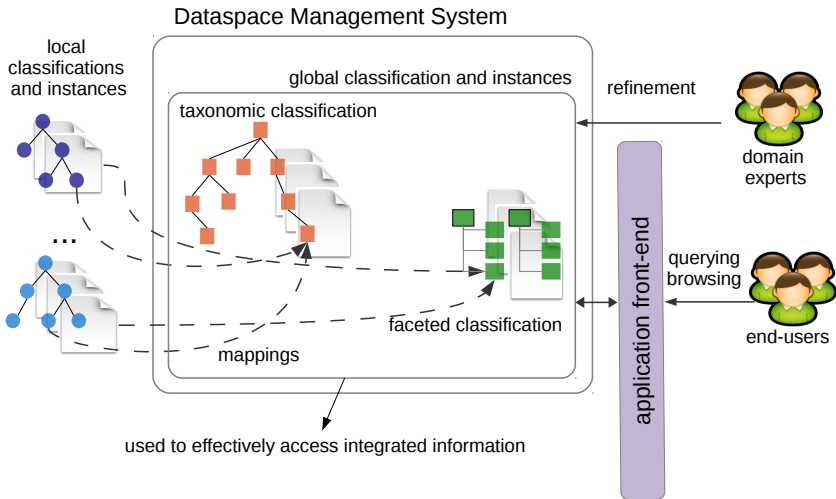
Quality-driven Evolution in Information Integration Systems

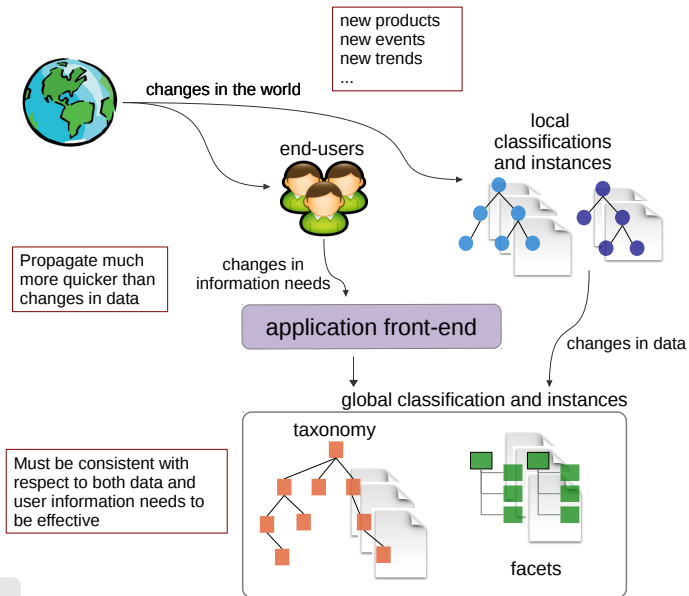
Ph.D. in Computer Science Course - XXVIII Series

Riccardo Porrini

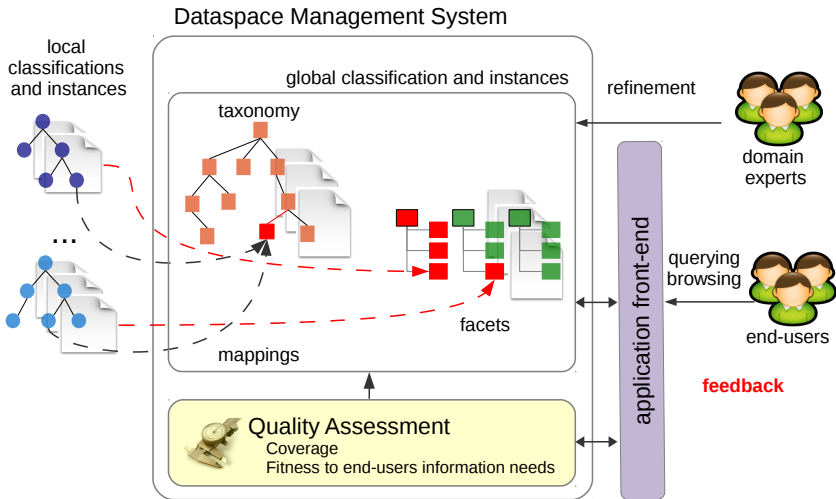
Supervisor: Dott. Palmonari

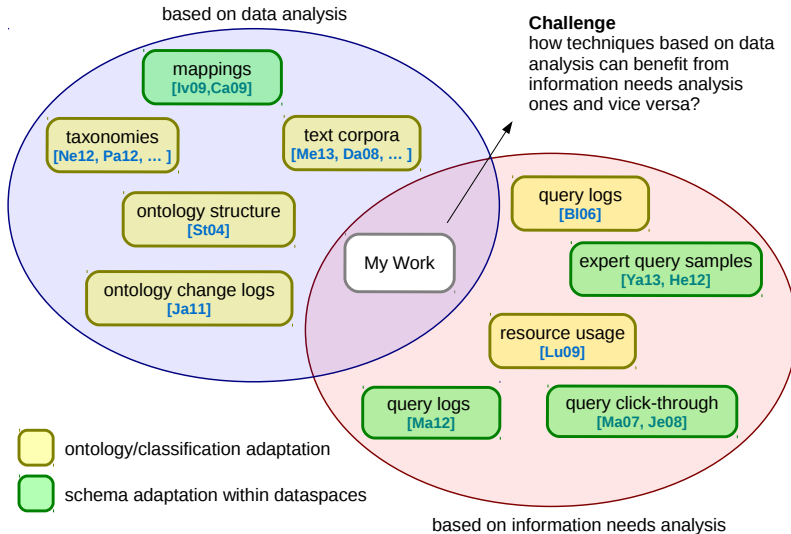
Tutor: Prof. Messina



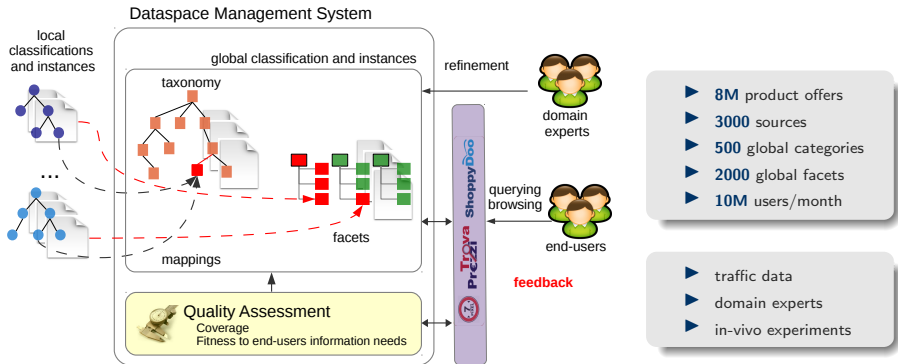


Goal: Quality-driven Information Integration Evolution



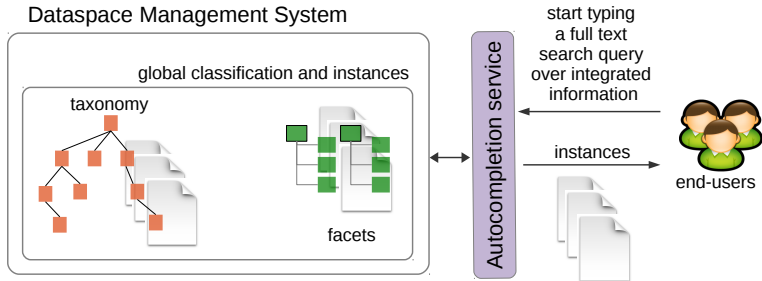


Evaluation using Real World Case Studies



Usage of global classifications to fulfill end-users' explorative information needs

Autocompletion algorithm that leverages both taxonomic and faceted classifications to suggest Dataspace instances [WI2012, WIAS2014]

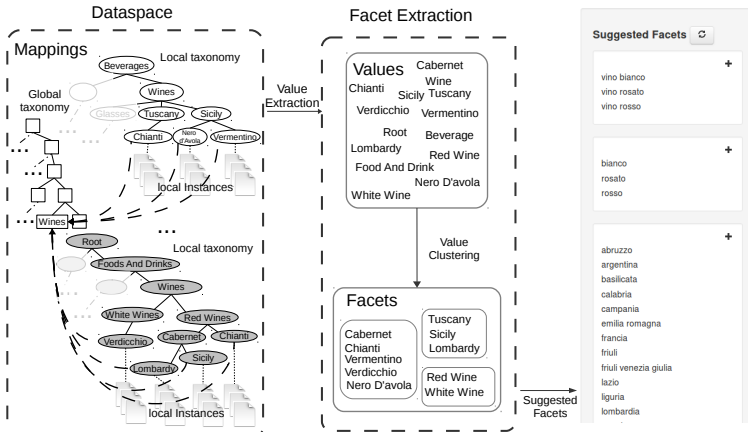


[WI2012] Palmonari et al. COMMA: A Result-Oriented Composite Autocompletion Method for E-marketplaces. In *Web Intelligence*, 2012

[WIAS2014] Porrini et al. Composite Match Autocompletion (COMMA): a Semantic Result-Oriented Autocompletion Technique for e-Marketplaces. In *Web Intelligence and Agent Systems Journal*, 2014

Adaptation of the global faceted classification based on data analysis

Extraction of facets from heterogeneous source taxonomies [CAiSE2014]



[CAiSE2014] Porri et al. Extracting Facets from Lost Fine-Grained Classifications in Dataspace. In CAiSE, 2014

Accepted Facets



Contributions

- ▶ Integration of heterogeneous taxonomies at Web scale
- ▶ **Taxonomy Layer Distance**: a new distance metric between heterogeneous taxonomic concepts
- ▶ **Evaluation** using real world data and comparison with existing metrics
- ▶ Currently deployed and used by TrovaPrezzi domain experts

Suggested Facets



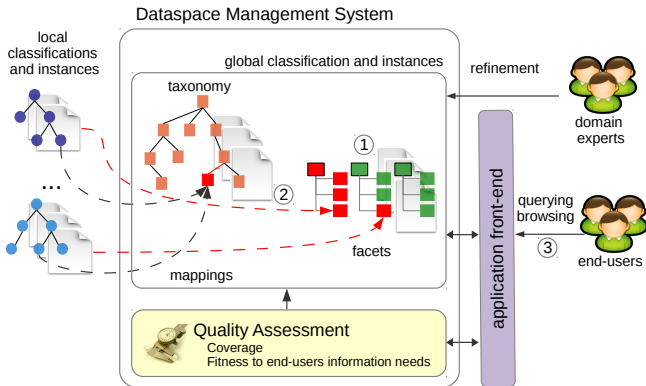
vino bianco
vino rosato
vino rosso



bianco
rosato
rosso



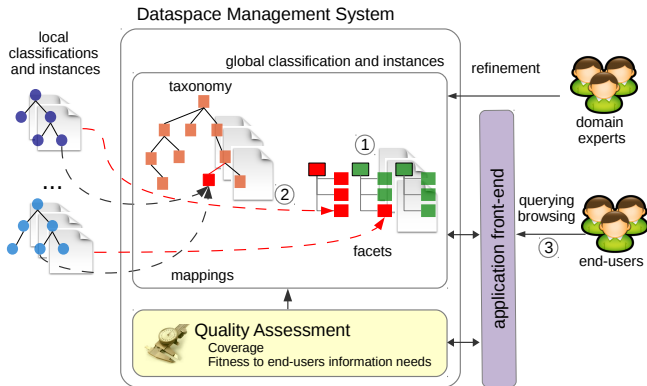
abruzzo
argentina
basilicata
calabria
campania
emilia romagna
francia
friuli
friuli venezia giulia
lazio
liguria
lombardia



①

Reconcile extracted facets to shared vocabularies (e.g., DBpedia ontology) (@Advis Lab - UIC - Chicago)

Adaptation from data analysis



①

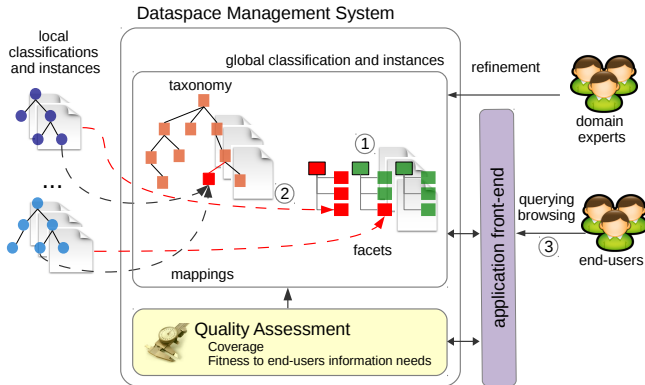
Reconcile extracted facets to shared vocabularies (e.g., DBPedia ontology) (@Advis Lab - UIC - Chicago)

Adaptation from data analysis

②

Suggest effective classification changes by mining search logs

Adaptation from information needs analysis



①

Reconcile extracted facets to shared vocabularies (e.g., DBPedia ontology) (@Advis Lab - UIC - Chicago)

Adaptation from data analysis

②

Suggest effective classification changes by mining search logs

Adaptation from information needs analysis

③ (side activity)

Study of the correlation between experimental economic indicators and user search behavior mined by search logs

Correlating dynamic information needs with "external" factors

Publications

- [1] Porrini et al. Extracting Facets from Lost Fine-Grained Classifications in Dataspace. In *CAiSE*, 2014
- [2] Porrini et al. Composite Match Autocompletion (COMMA): a Semantic Result-Oriented Autocompletion Technique for e-Marketplaces. In *Web Intelligence and Agent Systems Journal*, 2014
- [3] Palmonari et al. Comma: A Result-Oriented Composite Autocompletion Method for e-Marketplaces. In *Web Intelligence*, 2012

Conference Paper Reviews (as sub-reviewer)

- ▶ **AAAI-15**: 23th AAAI Conference on Artificial Intelligence
- ▶ **EKAW 2014**: 19th International Conference on Knowledge Engineering and Knowledge Management
- ▶ **ISWC 2014**: 14th International Semantic Web Conference
- ▶ **WI 2014**: 2014 IEEE/WIC/ACM International Conference on Web Intelligence
- ▶ **AAAI-14**: 22th AAAI Conference on Artificial Intelligence
- ▶ **ESWC 2014**: 11th Extended Semantic Web Conference
- ▶ **CAISE 2014**: 26th International Conference on Advanced Information Systems Engineering
- ▶ **ODBASE 2013**: 12th International Conference on Ontologies, DataBases, and Applications of Semantics
- ▶ **WI 2013**: 2013 IEEE/WIC/ACM International Conference on Web Intelligence

Courses and Schools

- ▶ The Impact of Logic: from Proof Systems to Databases - Politecnico di Milano (*evaluation pending*)
- ▶ Cluster Analysis - DISCo
- ▶ Third ESWC Summer School - Kalamaky - Crete (GR)
- ▶ Foundations of Data Exchange and Integration - Politecnico di Milano
- ▶ Advanced Analytics and Behavior Informatics - DISCo

Seminars

- ▶ From Sentiment Analysis to Continuous Learning - DISCo
- ▶ Introduzione alla Logica nella Rappresentazione della Conoscenza - DISCo
- ▶ La Misurazione Della Felicità al Tempo dei Big Data - DISCo
- ▶ Optimum Hyperpaths in Directed Hypergraphs - DISCo
- ▶ Phase Transitions in Social and Economic Systems - DISCo
- ▶ Progettare e Fare Open Data. Metodologia e tools sviluppati in Evodevo a partire dall'esperienza Open Data INPS - DISCo
- ▶ Semantic Constraints for Data Quality Assessment and Cleaning - DISCo
- ▶ Big Data e la forza degli eventi - DISCo
- ▶ WOA 2012: 13th National Workshop "Dagli Oggetti agli Agenti" - DISCo

Teaching

- ▶ Tutor for *Data and Web Semantics* course (MSc and PhD, Fall 2014) - UIC
- ▶ Co-Advisor of two student BSc thesis
- ▶ Lecturer (2 lectures) for the "*Artificial Intelligence*" course (MSc A.A. 2012/2013 and 2013/2014) - DISCo
- ▶ Tutor for *Distributed Systems* course (BSc A.A. 2012/2013) - DISCo

Questions?