Chapter 13

Related Work

This chapter compares the WIQA framework with related work.

Database Views. Accepted graphs within the WIQA framework can be compared to views in the context of relational databases. WIQA's explanation capabilities relate to work within the database community on explaining data lineage and view generation. An approach to explaining view generation has been developed by Cui and Wisdom in the context of the Stanford University WHIPS data warehousing system [CW00]. For a given data item in a materialized view, the authors propose a lineage tracing algorithm to identify the set of source data items that produced the view item. The algorithm is applicable to aggregate-select-project-join views and can be employed by data warehouse analysis tools to provide drill-down functionality from view items to source data items. What distinguishes the WIQA framework from the work within the relational data base community is the underlying data model. By employing a variation of the RDF data model, the WIQA framework is tailored towards the integration of heterogeneous information from the Web. For instance, integrating two partial descriptions of the same object while keeping track of the provenance of different pieces of information can simply be achieved within the Named Graphs data model, but is tricky using the relational data model.

Inference Web. A related approach to explaining information quality in the context of web-based information systems has been developed by the Inference Web project at the Stanford University Knowledge Systems Laboratory [MdS03]. The project aims at making query answers more transparent by providing explanations about information sources as well as inference processes that are used to derive query results. The Inference Web infrastructure includes a registry containing details on

information sources, reasoners, languages, and rewrite rules; a portable proof specification language; and a proof and explanation browser. Inference Web and the WIQA framework assume different application scenarios. While the WIQA framework is tailored towards a simple web-based information integration scenario, Inference Web assumes an agent community consisting of actively reasoning agents that cooperatively derive query answers from shared knowledge. Therefore, Inference Web focuses on explaining distributed reasoning paths [McG96], while the WIQA framework generates explanations why subjective information filtering policies are satisfied.

TRELLIS. A further system that employs the RDF data model and provides information quality assessment functionality is the TRELLIS information analysis tool [GR02] developed at the University of Southern California. TRELLIS aims at supporting intelligence analysts in selecting quality information within a military setting. As an analyst makes a decision on which sources to dismiss and which to believe, TRELLIS captures the derivation of the decision in a semantic markup. The system then uses these annotations to derive an information quality assessment of the source based on the annotations of many individuals. Compared with the WIQA framework, TRELLIS supports only one fixed ontology for capturing quality-related meta-information and a single hard-coded assessment policy, while the WIQA framework may employ arbitrary ontologies and a wide range of different filtering policies.

Almendra and Schwabe. An approach for translating abstract information quality requirements into concrete filtering policies is presented by Almendra and Schwabe from the Pontificia Universidade Catolica do Rio de Janeiro in [AS05]. Their work is directly based on the work presented in this thesis and employs the Named Graphs data model, the Semantic Web Publishing Vocabulary and the TriQL.P policy language [Biz04b], an earlier version of the WIQA-PL policy language. In addition to this work, where each policy must explicitly specify all the conditions that triples must fulfill to be accepted, they propose an ontology for expressing information quality requirements (TrustPoints) and an algorithm that automatically derives information filtering policies from these requirements by combining query fragments. Given adequate tool support, their translation mechanism provides a valuable extension to this work as it reduces the technical knowledge required from a policy author.