

Im Rahmen des SemIPort-Projektes wurde für die Bereiche Browsing und Visualisierung ein Mensch-Maschine-Interface entwickelt, welches dem Benutzer bei seiner Suche nach den gewünschten Informationen unterstützt und leitet (siehe Abbildung 2). Der DBL-Browser ermöglicht durch die vollkommene Verlinkung der bibliographischen Daten ein Navigieren durch den gesamten Datenbestand und ist ähnlich einfach zu bedienen wie bekannte Internet-Browser. Durch die IR-Schnittstelle sind auch komplexere Suchen möglich. Der DBL-Editor bietet die Möglichkeit weitere Datensätze auf hohem Qualitätsniveau manuell zu erfassen und dabei „on-the-fly“ nach bereits bekannten Daten zu suchen, um Duplikate schon bei der Eingabe zu vermeiden [KLRRWW04].

Neben den Daten des DBLP-Servers, welche in den Hauptspeicher des Computers eines Benutzers geladen werden können, besteht die Möglichkeit über das Internet die Daten des **io-port.net** (siehe Abschnitt 1) mit Hilfe von SWQL (siehe Abschnitt 10) zu browsen und nach Autoren sowie nach Wörtern im Titel einer Publikation zu suchen.

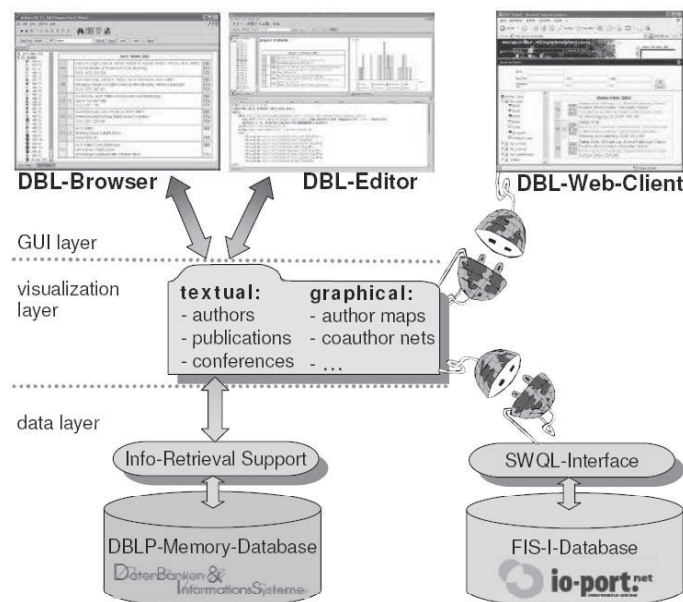


Abbildung 2: Schematischer Aufbau der DBL-Komponenten

2.4 SWQL – Eine OWL-basierte Abfragesprache zur Datenintegration

Zahlreiche Anwendungen besonders aus dem „Semantic Web“ Umfeld basieren auf einer Ontologie, welche die Domäne der Anwendung modelliert. Die Mehrzahl der existierenden Datenquellen beziehen sich jedoch nicht direkt auf eine Ontologie, sondern sind z.B. als Relationen oder in XML modelliert. Diese Datenmodelle haben ihre eigenen Schema- und Abfragesprachen, wie ein relationales Schema und SQL bzw. XML Schema und XQuery. Um solche Datenquellen über eine Ontologie abzufragen, wird

eine Abfragesprache benötigt, die in der Lage ist, Abfragen an Ontologie-basierte Daten zu stellen, d.h., Instanzen eines Konzepts auszuwählen und aufgrund von Attributwerten zu filtern, zwischen Instanzen über Beziehungen zu navigieren sowie Attributwerte auszuwählen. Hierzu wurde die Abfragesprache Semantic Web Query Language (SWQL) entwickelt [LSF04]. Sie basiert auf der Syntax der derzeit vom W3C standardisierten XML-Abfragesprache XQuery. Allerdings wurde die Semantik, das Typsystem und das Datenmodell von XQuery vollständig ersetzt.

SWQL nutzt die Web Ontologie Sprache OWL als ihr Typsystem, d.h. eine Abfrage wird mit Hilfe des Vokabulars einer Ontologie formuliert und kann gegen diese auch geprüft werden. Das zugehörige Datenmodell für SWQL ist ein Graph-basiertes Datenmodell, das ähnlich dem RDF-Datenmodell ist. Außerdem wurde XPath, die Sprache in XQuery zur Selektion einzelner Knoten im XQuery-Datenmodell, durch SWQLPath, einer Sprache zur Selektionen einzelner Knoten in einem SWQL-Graphen, ersetzt. Im Gegensatz zu existierenden RDF-Abfragesprachen, die von einem ähnlichen Datenmodell und Typsystem ausgehen, ist SWQL eine vollständige Abfragesprache, die neben Selektion und Navigation auch die Konstruktion neuer Datenstrukturen erlaubt. Dies erlaubt es, Datenstrukturen zu ändern, anzupassen und zu übersetzen, was Voraussetzung zum Einsatz als Datenintegrationssprache ist.

SWQL ist außerdem eine streng getypte Sprache, was eine statische Prüfung von SWQL-Abfragen auf Fehler erlaubt. So kann beispielsweise die Kompatibilität der Parameter von Funktionsaufrufen oder der Typen bei der Konstruktion von Konzepten und Beziehungen bereits vor der Ausführung der Abfrage geprüft werden. Dies ist besonders dann nützlich, wenn Abfragen nicht direkt ausprobiert werden können, z.B. im Fall von Funktionsbibliotheken oder sehr lang dauernden Abfragen.

Um SWQL-Abfragen gegen verschiedene reale Datenquellen wie XML-, RDF- oder Relationale Datenbanken stellen zu können erwies sich eine Schichtenarchitektur als sehr flexibel und performant, dabei wird für jeden Datenquellentyp ein Übersetzungsmodul für das Datenmodell sowie eine Funktionsbibliothek zum Datenbankzugriff geschrieben. Diese Architektur erlaubte es, innerhalb kürzester Zeit SWQL-Abfragen an eine XML- und eine RDF-Datenbank sowie gegen die von FIS-I eingesetzte Lucene-Datenstruktur auszuführen.

Für die Integration konkreter Datenquellen in ein Ontologie-basiertes Portal müssen u.a. die Schemata der einzelnen Datenquellen integriert werden, dazu werden SWQL-Funktionsbibliotheken genutzt. Diese Funktionsbibliotheken bestehen aus feingranularen Funktionen zur Übersetzung einzelner Konzepte aus dem lokalen Schema in die globale Ontologie. Die vollständigen Funktionsbibliotheken können mit Hilfen der statischen Typüberprüfung garantieren, nur korrekte Instanzen der Zielontologie zu erzeugen. Als lokale Schemata können neben OWL-Ontologien auch XML-Schemata dienen [LF04]. Diese Schemaintegrationsmethode lässt sich sowohl für das Füllen eines Datawarehouses, als auch für eine Mediator-Wrapper-Architektur einsetzen.