Zusammenfassung

Die Integration und Beschreibung großer Mengen von unterschiedlich stark strukturierten Daten ist eine aktuelle Herausforderung für die Datenbankwelt. In vielen Bereichen werden lediglich speziell zugeschnittene ad-hoc Lösungen verwendet. Es fehlt an allgemeinen Konzepten, welche den universellen Einsatz und die Wiederverwendbarkeit einmal entworfener Lösungen gewährleisten.

Bei der Integration und Beschreibung der anfallenden Daten fällt auf, dass diese oftmals eine inhärente Struktur besitzen. Daher liegt es nahe, diese Struktur entsprechend zu verwalten. Da die auftretenden Strukturen in der Regel tief geschachtelt, unregelmäßig oder auch implizit vorhanden sind, erweisen sich konventionelle Datenmodelle für deren Beschreibung als unzureichend. Als geeignete Alternative stellt sich die Extensible Markup Language (XML) dar, die zunächst im Internet als globale Datenaustauschsprache ihren Ursprung hatte.

Im Rahmen dieser Arbeit werden XML Document Type Definitions (DTDs) bezüglich ihrer formalen Eigenschaften untersucht. Ausgehend von dieser Analyse wird eine Erweiterung definiert, die DTDs zur Beschreibung von Schemata tauglich machen. Die vorgestellte Analyse und Erweiterung kann über diese Arbeit hinaus als Grundlage für XML-basierte Modelle dienen. Ausgehend von der XML-Erweiterung wird ein Modell zur Beschreibung von Objekten, Objektstrukturen und -repositorien entwickelt. Bei der praktischen Umsetzung des Modells werden neben Anfrageoperationen auch Migrations- und Restrukturierungsmöglichkeiten betrachtet.

Die Arbeit mündet in der Erstellung eines Prototypen, der zur Unterstützung von Forschung und Lehre an der Medizinischen Universität zu Lübeck einem breiten Anwenderfeld zu Gute kommt. Dieser Prototyp verifiziert die theoretischen Ergebnisse dieser Arbeit und zeigt deutlich den praktischen Nutzen auf, den die entwickelten Konzepte dem Benutzer liefern. Es stellt sich heraus, dass XML in Kombination mit angrenzenden Standards als Basis für die Strukturierung, Klassifikation und inhaltliche Beschreibung im Rahmen eines Medienarchivs eine wesentliche Verbesserung und Flexibilisierung gegenüber herkömmlichen Datenmodellen darstellt.

Abstract (english)

The integration and description of vast amounts of semistructured data presents a real challenge to the database community. In general only proprietary solutions are applied. There is a lack of common concepts which would provide the universal deployment and reusability of once defined solutions.

We can assume that the information usually bears an inherent structure. Therefore, it suggests itself to try to manage this information efficiently. Since the hierarchical structures are often deep, irregular or implicit, conventional data models seem to be insufficient for this problem. The Extensible Markup Language (XML), which was first used in the internet area, seems to be a promising alternative.

In this thesis XML Document Type Definitions (DTDs) are investigated closely, specifically focusing on their formal properties. Based on this analysis an extension is proposed, making DTDs more appropriate for schema definition. The presented analysis and extension can be used beyond this thesis as a formal framework for XML-based modeling. Starting from this extension, a model for describing objects, object structures and repositories is designed. Beyond the practical implementation, searching and restructuring issues are considered.

This thesis leads to the development of a prototype, which is used for research and lectures in the Medical University of Luebeck. The prototype verifies the theoretical results presented here and shows clearly the benefit of the new concepts provided to the user. It is shown that, in contrast to conventional models, the combination of XML with other standards greatly improves the structuring, classification and textual description in the context of a media archive.