Querying Sequence Databases with Transducers

Anthony Bonner
Department of Computer Science
University of Toronto — Canada M5S 1A1
bonner@db.toronto.edu

Giansalvatore Mecca
D.I.F.A. — Università della Basilicata
Via della Tecnica, 3 — 85100 Potenza, Italy
mecca@dis.uniroma1.it

Abstract

This paper develops a database query language called *Transducer Datalog* motivated by the needs of a new and emerging class of database applications. In these applications, such as text databases and genome databases, the storage and manipulation of long character sequences is a crucial feature. The issues involved in managing this kind of data are not addressed by traditional database systems, either in theory or in practice. To address these issues, in recent work, we introduced a new machine model called a *generalized sequence transducer.* These generalized transducers extend ordinary transducers by allowing them to invoke other transducers as "subroutines." This paper establishes the computational properties of Transducer Datalog, a query language based on this new machine model. In the process, we develop a hierarchy of time-complexity classes based on the Ackermann function. The lower levels of this hierarchy correspond to well-known complexity classes, such as polynomial time and hyper-exponential time. We establish a tight relationship between levels in this hierarchy and the depth of subroutine calls within Transducer Datalog programs. Finally, we show that Transducer Datalog programs of arbitrary depth express exactly the sequence functions computable in primitive-recursive time.

Appears in *Proceedings of the Sixth International Workshop on Database Programming Languages (DBPL)*, August 18–20, 1997, Estes Park, Colorado.