# Finite Query Languages for Sequence Databases

**Giansalvatore Mecca***

Universitá di Roma "La Sapienza"
Dipartimento di Informatica e Sistemistica
Via Salaria, 113 — 00198 Roma, Italy

`mecca@infokit.dis.uniroma1.it`

**Anthony J. Bonner†**

University of Toronto
Department of Computer Science
Toronto, Ontario — Canada M5S 1A4

`bonner@db.toronto.edu`

**Abstract**

This paper develops a query language for sequence databases, such as genome databases and text databases. Unlike relational data, queries over sequential data can easily produce infinite answer sets, since the universe of sequences is infinite, even for a finite alphabet. The challenge is to develop query languages that are both highly expressive and finite. This paper develops such a language. It is a subset of a recently developed logic called *Sequence Datalog* [22]. Sequence Datalog distinguishes syntactically between *subsequence extraction* and *sequence construction*. Extraction creates sequences of bounded length, and leads to *safe* recursion; while construction can create sequences of arbitrary length, and leads to *unsafe* recursion. In this paper, we develop syntactic restrictions for *Sequence Datalog* that allow sequence construction but preserve finiteness. The main idea is to use safe recursion to control and limit unsafe recursion. The main results are a finite language, called *Weakly Constructive Sequence Datalog*, and a characterization of its complexity and expressive power. Although finite, the new language is highly expressive, since its data complexity is complete for the elementary functions.

Appears in *Proceedings of the Fifth International Workshop on Database Programming Languages (DBPL)*, September 6–8, 1995, Gubbio, Umbria, Italy.

This and related papers are available at the following URL:
`http://www.cs.toronto.edu/~bonner/papers.html#sequence-datalog`

*Research partially supported by MURST and Consiglio Nazionale delle Ricerche (CNR).

†Research partially supported by an operating grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).