2009

Knowledge libraries and information space

Eric Rayner
University of Wollongong

KNOWLEDGE LIBRARIES AND INFORMATION SPACE

A thesis submitted in partial fulfilment of the requirements for the award of the degree

DOCTOR OF PHILOSOPHY

from the

UNIVERSITY OF WOLLONGONG

by

ERIC RAYNER, BCompSc(Hons)

SCHOOL OF COMPUTER SCIENCE AND SOFTWARE ENGINEERING

2009

Thesis Certification

I, Eric P. Rayner, declare that this thesis, submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Information and Computer Science, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Eric Rayner
July 27, 2009

Contents

List of Figures

List of Tables

Abbreviation, Notation and Typographical Conventions

The set of real numbers is denoted by $\mathbb{R}$, the set of rational numbers by $\mathbb{Q}$ and the set of natural numbers (integers, strictly larger than 0) by $\mathbb{N}_1$. Note that $\mathbb{R}^{\geq 0}$ is the set of real numbers greater than or equal to 0, while $\mathbb{R}^{>0}$ is the set of real numbers strictly greater than 0. Interval notation is used to denote real intervals, so $(0, 1]$ is the set of real numbers less than or equal to 1 and strictly larger than 0. More generally, capital letters, such as L, M, X, Y, are used to denote sets. The power set of any set M (the set of all subsets of M) is denoted $\mathcal{P}(M)$.

Lowercase "math bold font" letters denote vectors, so $\mathbf{x}$ and $\mathbf{y}$ are vectors.

L-collections (introduced in chapter 6) are distinguished from sets by using "math calligraphy font", so $\mathcal{M}, \mathcal{X}, \mathcal{Y}$ are L-collections.

Enclosing vertical bars are used to denote the cardinality of a set ($|M|$), the cardinality of an L-collection ($|\mathcal{M}|$), the absolute value of a real number ($|d(x, y)|$) and the magnitude of a vector ($|\mathbf{x}|$).

Bold text is used to denote key terms that are defined (or at least described), both within, and (optionally) prior to, the definition. "Double quotes" are used for short quotations (which are also referenced) and when introducing key terms that are not defined.

Finally, *iff* is used as shorthand for "if and only if".

Abbreviations in this thesis are preceded and introduced by the corresponding, non abbreviated, full term.

Abstract

This research describes and develops **Knowledge Libraries**, idealised systems for organising and presenting information. By providing a mathematical basis, the definition of **information space** establishes a formal foundation for Knowledge Libraries. The definition of information space builds on the new definitions of L-**collections**, which generalise sets by allowing a real valued grade to be associated with each element, and **set space**, which generalises metric space to better model the relationships between **information units**.

The **multiple search tree** method improves existing metric space range query algorithms. These algorithms are also generalised to work over set space. The **sequential-hybrid algorithm** enables efficient range queries over multi-dimensional spaces.

Acknowledgements

First and foremost, I would like to thank Dr. Ian Piper, my principal supervisor, for his continuing support throughout my candidacy. Acknowledgement is also due Prof. Martin Bunder for his assistance with the mathematical work in this thesis. Without Prof. Bunder's input, this thesis would not have the emphasis on mathematical correctness that it does.

I would also like to thank my family for their support, particularly my father, Prof. John Rayner, for his willingness to proof read chapters and discuss thesis related matters (at some considerable length).