



Università degli Studi di Salerno

Dipartimento di Informatica

Dottorato di Ricerca in Informatica
Curriculum Computer Science and Information
Technology
XXXVIII Ciclo

TESI DI DOTTORATO / PH.D. THESIS

From Search to Coding (and Back): Two Sides of the Same Coin

ROBERTO BRUNO

SUPERVISOR:

PROF. UGO VACCARO

PHD PROGRAM DIRECTOR:

PROF. ANDREA DE LUCIA

A.A 2024/2025

ABSTRACT

This thesis investigates the deep connection between deterministic search procedures and *variable-length codes* (VLCs), showing that these two domains—traditionally treated as distinct—are fundamentally equivalent. Any search algorithm based on membership or comparison queries naturally induces a binary *variable-length code* (VLC), and conversely, every VLC implicitly represents a search strategy. This correspondence allows classical search problems to be reformulated through the mathematical language of Coding Theory and analyzed using tools from Information Theory.

Within this framework, the thesis investigates several families of alphabetic and prefix codes arising from different structural or cost constraints. First, it establishes new and tighter upper bounds for optimal alphabetic codes and introduces linear-time algorithms that achieve them, yielding improved bounds for *binary search trees* (BSTs) as well. Second, it proposes and analyzes a novel class of codes, denoted as (α, β) -constrained codes, motivated by search problems with asymmetric test outcome costs, and provides polynomial-time algorithms for their construction. Third, it studies prefix codes that use a symbol only as a termination character, deriving linear-time near-optimal construction methods and new entropic bounds. Moreover, it explicitly connects these codes to search procedures using comparison and equality tests.

Overall, the thesis shows that search can be understood as an act of compression: efficient information acquisition and efficient information representation are governed by the same principle of minimizing uncertainty. This perspective provides a powerful, general methodology for translating search problems into code problems.

SOMMARIO

Questa tesi indaga la profonda connessione tra le procedure di ricerca deterministiche e i codici a lunghezza variabile, dimostrando come questi due ambiti — apparentemente distinti — siano in realtà equivalenti. Ogni algoritmo di ricerca basato su test di appartenenza o di confronto induce naturalmente un codice a lunghezza variabile e, viceversa, ogni codice a lunghezza variabile rappresenta implicitamente una strategia di ricerca. Tale corrispondenza consente di riformulare i problemi classici di ricerca nel linguaggio matematico della Teoria dei Codici e di analizzarli mediante strumenti della Teoria dell'Informazione.

All'interno di questo quadro, la tesi analizza diverse famiglie di codici alfabetici e prefissi, caratterizzate da differenti vincoli strutturali o di costo. In primo luogo, stabilisce nuove limitazioni superiori sulla lunghezza media di codici alfabetici ottimi ed introduce algoritmi lineari che ottengono tali limitazioni, usandoli poi anche per derivare nuove limitazioni anche sulla lunghezza media di Alberi Binari di Ricerca. In secondo luogo, propone e analizza una nuova classe di codici, motivata da problemi di ricerca con costi asimmetrici degli esiti dei test, e fornisce algoritmi polinomiali per la loro costruzione. In terzo luogo, esamina codici prefisso che utilizzano un simbolo esclusivamente come carattere di terminazione, derivando metodi di costruzione quasi ottimali in tempo lineare ed anche nuove limitazioni sulla loro lunghezza in termini di entropia. Inoltre, mette in evidenza l'esplicita connessione tra questa nuova classe di codici e le procedure di ricerca che impiegano test di confronto ed eguaglianza.

Nel complesso, la tesi mostra come la ricerca possa essere interpretata come un processo di compressione: l'efficienza nell'acquisizione dell'informazione e l'efficienza nella sua rappresentazione sono governate dallo stesso principio di minimizzazione dell'incertezza. Questa prospettiva fornisce una metodologia generale e potente per tradurre problemi di ricerca in problemi di costruzione di codici.