

Povzetek

Ta diplomska naloga obravnava izbrane teme iz teorije abstraktnih avtomatov in formalnih jezikov. Predstavljena je postopna rešitev problema iskanja vzorcev besed v besedilih na osnovi prepoznavanja in tolmačenja regularnih izrazov s poudarkom na teoretični podlagi. Obravnavane so teme iz formalnih jezikov, slovnice s frazno zgradbo, regularnih množic in regularnih izrazov, izračunljivih funkcij, končnih avtomatov in njihovih jezikov.

Prvo poglavje je posvečeno izgotovitvi uporabne matematične definicije celotnega problema, kar vključuje osnovne definicije iz algebre besed, odkrivanje dosegljivih, uporabnih pogojev za določanje oblike vzorcev v besedilu in vpeljave primerne matematičnega izrazja za opisovanje pogojev za določanje oblik besed (tu se spoznamo s pojmom regularnih množic in regularnih izrazov), kar umestimo tudi v širši okvir hierarhije formalnih jezikov, hierarhije Chomskega in za regularne izraze posebej med kontekstno neodvisne jezike. Bralec, seznanjen s temi pojmi, lahko poglavje brez škode za razumevanje v nadaljevanju izpusti. Poglavje se zaključi z natančno matematično definicijo problema, ki ga v nadaljevanju rešimo.

Drugo poglavje je namenjeno izbiri modela za rešitev problema, kot je definiran na koncu prvega poglavja. Zamislimo si napravo, ki bi ustrezala potrebam rešitve problema tako, da njeno delovanje predstavimo z vhodno-izhodno funkcijo in za to funkcijo zahtevamo nekaj lastnosti; izkaže se, da lahko problem rešimo s sekvencično obdelavo in s končno prostorsko zahtevnostjo predstavitev v računalniku. Ugotovljene lastnosti opišemo z modeli teorije končnih avtomatov. Predstavimo deterministične končne avtomate, nedeterministične končne avtomate in nedeterministične končne avtomate z ε -prehodi. Vse te matematične modele potrebujemo, da lahko preidemo od prvotnega determinističnega modela, ki ustreza ugotovitvam z začetka poglavja, do učinkovitega modela, ki ga uporabimo za končno rešitev problema.

V tretjem poglavju se ukvarjamo z jeziki končnih avtomatov in ovrednotimo modele končnih avtomatov, predstavljene v drugem poglavju, s stališča prepoznavanja teh jezikov. Ugotovimo, da so si modeli enakovredni. Spotoma za vsak model končnega avtomata vpeljemo še pojem sprejemnika kot okrnjenega avtomata, ki zadošča za rešitev problema prepoznavanja jezikov in ki ga bomo dejansko uporabili za rešitev problema. Upravičimo našo rabo končnega nedeterminističnega sprejemnika z ε -prehodi za rešitev problema. Poglavje zaključimo z dokazom, da so regularne množice natanko jeziki, ki jih lahko prepoznajo končni avtomati in s tem dokažemo, da so izbrani modeli res pravo sredstvo za rešitev našega problema.

V četrtem poglavju predstavimo pristop za rešitev problema, ki je primeren za kasnejšo računalniško implementacijo. Za rešitev problema izberemo model končnega nedeterminističnega sprejemnika z ε -prehodi in pokažemo, kako se da postopno sestaviti takšen sprejemnik za poljubno regularno množico, ki jo označuje ustrezen regularni izraz. V nadaljevanju izdelamo predstavitev sprejemnika z ustrezno podatkovno strukturo, primerno za implementacijo v računalniku in izgotovimo algoritem za simulacijo delovanja sprejemnika. Poglavje zaključimo z

izdelavo postopkov za izgradnjo sprejemnika za jezik regularne množice na osnovi prepoznavanja regularnega izraza, ki to množico označuje.

Sledi še dodatek, v katerem je opisana vsebina priložene diskete. Izdelana je knjižnica s funkcijami za izdelavo sprejemnikov na osnovi danih regularnih izrazov in za simulacijo sprejemnikov v namen razpoznavanja vzorcev besed v besedilih.

Math. Subj. Class. (2000): 68Q42, 68Q45, 03D40

Ključne besede: regularne množice, regularni izrazi,
formalni jeziki, končni avtomati

Keywords: regular sets, regular expressions,
formal languages, finite automata

Literatura

- [1] Arbib, Michael A. : Theories of abstract automata
- [2] Fu, K. S. : Syntactic Methods in Pattern Recognition
- [3] Ginsburg, S. : Algebraic and Automata-Theoretic Properties of Formal Languages
- [4] Hopcroft, John E. ; Ullman, Jeffrey D. : Uvod v teorijo avtomatov, jezikov in izračunov (slovenski prevod)
- [5] Sedgewick, Robert : Algorithms