

## KRATEK POVZETEK VSEBINE

V diplomskem delu so predstavljene mere podobnosti nizov. Prikazana so področja, kjer se mera podobnosti pogosto uporablja. Seznanimo se z elementarnimi operacijami, ki nam ponazarjajo osnovne razlike med nizi in s tremi različnimi predstavitvami analize različnosti med nizi, ki temeljijo na elementarnih operacijah. Te so pretvorba, poravnavanje in sled. Predstavljena je teorija abstraktne mere, s pomočjo katere lahko predstavimo mere podobnosti nizov, kot so : posplošena Levenshteinova razdalja, verjetnost napačnega niza, dolžina najdaljšega skupnega podniza, dolžina najkrajšega skupnega nadniza; pa tudi mere, katerih zaloga vrednosti so množice nizov : množica najdaljših skupnih podnizov, množica najkrajših skupnih nadnizov in množica zlitij. Na lastnostih abstraktne mere sloni izpeljava algoritma za izračun mere podobnosti. Vse našteje mere so izpeljane za par nizov in kasneje posplošene na končno množico nizov. Ob koncu so dodani še rezultati razvrščanja v skupine s programskim paketom CLUSE. Razvrstitve temeljijo na eni od zgoraj omenjenih mer izračunane s programom, ki je bil narejen v ta namen. Razvrščene so tri skupine podatkov: svetovni jeziki na osnovi besed istega pomena, sesalci glede na zobno formulo in vretenčarji na osnovi citokroma c.

Math. Subj. Class.(1985) : 62H20, 94A17, 68Q20,  
68T10, 62H30, 92-08

Key words : similarity measures between strings, transformations of strings, semirings, Levenshtein distance, cluster analysis, cytochrome - c, dentition of mammals, languages

# LITERATURA

- [1] Vladimir BATAGELJ : SIMILARITY MEASURES BETWEEN STRUCTURED OBJECTS.  
I.Proceedings of the MATH/CHEM/COMP 88, Dubrovnik.
- [2] J. A. HARTIGAN : CLUSTERING ALGORITHMS. John Wiley, London 1971.
- [3] Daniel S. HIRSCHBERG : ALGORITHMS FOR THE LONGEST COMMON SUBSEQUENCE PROBLEM.  
Journal of the Association for Computing Machinery Vol.24, No.4, October 1977 ( 664 - 675 ).
- [4] Daniel S. HIRSCHBERG : A LINEAR SPACE ALGORITHM FOR COMPUTING MAXIMAL COMMON SUBSEQUENCE. Communications of the ACM Vol.18, No. 6, June 1975.
- [5] James W. HUNT, Thomas G. SZYMANSKI : A FAST ALGORITHM FOR COMPUTING LONGEST COMMON SUBSEQUENCES. Communications of the ACM Vol.20, No.5, May 1977.
- [6] Peter KARLSON : BIOKEMIJA, DZS 1980.
- [7] R.L. KASHYAP, B.J. COMMEN : A COMMON BASIS FOR SIMILARITY MEASURES INVOLVING TWO STRINGS. Intern.J.Computer Math. Vol.13, 1983 ( 17 - 40 ).
- [8] R.L. KASHYAP, B.J. COMMEN : SIMILARITY MEASURES FOR SETS OF STRINGS.  
Intern.J.Computer Math. Vol.13, 1983 ( 95 - 104 ).
- [9] Joseph B. KRUSKAL : AN OVERVIEW OF SEQUENCE COMPARISON : TIME WARPS, STRING EDITS AND MACROMOLEKULES. SIAM Review Vol.25, No.2, April 1983.
- [10] Roy LOWRANCE, Robert A. WAGNER : AN EXTENSION OF THE STRING-TO-STRING CORRECTION PROBLEM.  
Journal of the Association for Computing Machinery Vol.22, No.2, April 1976 ( 177 - 183 ).
- [11] Miloje M. RAKOČEVIĆ : GENI MOLEKULI JEZIK. Beogradska knjiga.
- [12] Robert A. WAGNER, Michael J. FISCSHER : THE STRING-TO-STRING CORRECTION PROBLEM.  
Journal of the Association for Computing Machinery Vol.21, No.1, January 1974 ( 168 - 173 ).
- [13] THE CONCISE DICTIONARY OF 26 LANGUAGES IN SIMULTANEUS TRANSLATION.  
Compiled by P. M. Bergman. A Signet Book from New America Library.