

# Problem maksimalnog zajedničkog podstabla

**Mladen Dilparić**

# Opis problema

- Za datu kolekciju stabala naći stablo koje je podstablo svakog stabla kolekcije takvo da je veličina čvorova maksimalna.
- U engleskoj literaturi se najčešće pominje kao Maximum Common Subtree problem ili Largest Common Subtree problem.
- Primena ovog problema u biohemiji, bioinformatičari, varijacije problema primenu(maksimalno ponavljajuće stablo) nalaze i u detekciji plagijarizma
- Pripada klasi NP problema
- U radu razmatrana neobeležena stabla, sa neuredjenom listom dece(poredak nije bitan)
- Nije nadjena adekvatna literatura ili radovi na temu optimizacije ovog problema

# Algoritam grube sile

- Iz kolekcije stabala izaberemo stablo koje je najmanje (sa najmanjim brojem čvorova)
- Generišemo sva podstabla – rekurzivno za decu , a potom generišemo sve kombinacije podstabala dece sa korenom
- Sortiramo generisana podstabla opadajuće po veličini čvorova, i krenemo sa proverom izomorfnosti
- Provera da li je stablo podstablo drugog stabla je polinomijalna (Uradjena svodjenjem problema na problem uparivanja u bipartitnom grafu – Hopcroft Karp)
- Prvo nadjeno podstablo svakog stabla kolekcije je maksimalno podstablo kolekcije
- Eksponencijalna složenost

# Algoritam simuliranog kaljenja

- Generacija inicijalnog rešenja – nasumično generisano podstablo kolekcije, jedan čvor, minimalno stablo kolekcije
- Fitnes funkcija – broj čvorova stabla ako je zajedničko podstablo kolekcije, 0 inače
- Generisanje okoline rešenja – zavisi od odabranog načina generisanja inicijalnog rešenja
- Parametri algoritma simuliranog kaljenja – različite temperature kaljenja probane

# Rezultati

- Korišćene test kolekcije su generisane pravljjenjem nasumično generisanih stabala sa odredjenim uslovljenostima(broj stabala dece po čvorovima, visina stabla). Korišćena binarna Knutova reprezentacija stabala.
- Rezultati izvršavanja algoritma simuliranog kaljenja su dati kao najbolji rezultat iz višestrukog pokretanja algoritma(do 5)
- Prilikom eksperimentalne analize najbolje rezultate je davalo generisanje nasumičnog stabla kolekcije

# Poredjenje rezultata

Test primer	Broj stabala	BF algoritam	SK algoritam	BF vreme	SK vreme
test1	9	20	20	1.025s	1.161s
test2	10	25	22	1min	3s
test12	14	18	18	2s	3s
test3	18	30	30	11s	11s
test19 10 4	19	25	25	13s	14s
test4	20	28	27	30min	8s
test40	40	-	70	-	204s
multi	15	13	15	250s	16s
mnogo	20	14	14	126s	3s
test49 8 5	49	28	25	400s	27s



# Zaključak

- U slučajevima sa manjim brojem čvorova, algoritam grube sile se pokazao kao bolji jer daje egzaktno rešenje u otprilike istom vremenu.
- Algoritam simuliranog kaljenja daje zadovoljavajuće rezultate na srednjim i većim primerima (visina stabla preko 5, broj čvorova po detetu većeg obima). Veličina nadjenog podstabla je uglavnom na test primerima varirala do 10-12 posto u odnosu na optimalna rešenja.
- Moguće je preprocesiranje kolekcije zarad dodatnog poboljšanja, razmatranje druge fitnes funkcije, drugačije konstrukcije inicijalnog rešenja, kao i generisanja okoline rešenja

# Literatura

- 1. Beleske sa predavanja iz Računarske Inteligencije
- 2. Knjiga Algorithms on Trees and Graphs by Gabriel Valiente
- 3. Akutsu, T., and Halldórsson, M. (1994), "On the approximation of largest common point sets and largest common subtrees"