

## Графови [100 поена]

Граф је нелинеарна структура података која се састоји од скупа чворова и скупа грана. Гране представљају односе (везе) између чворова. Графови се могу користити за моделирање произвољних нелинеарних релација. Постоје усмерени и неусмерени графови.

### Репрезентација графа

У зависности од редног броја  $i$  добијеног коришћењем формуле назначене у напоменама, потребно је користити једну од следећих меморијских репрезентација графа приликом решавања задатих проблема:

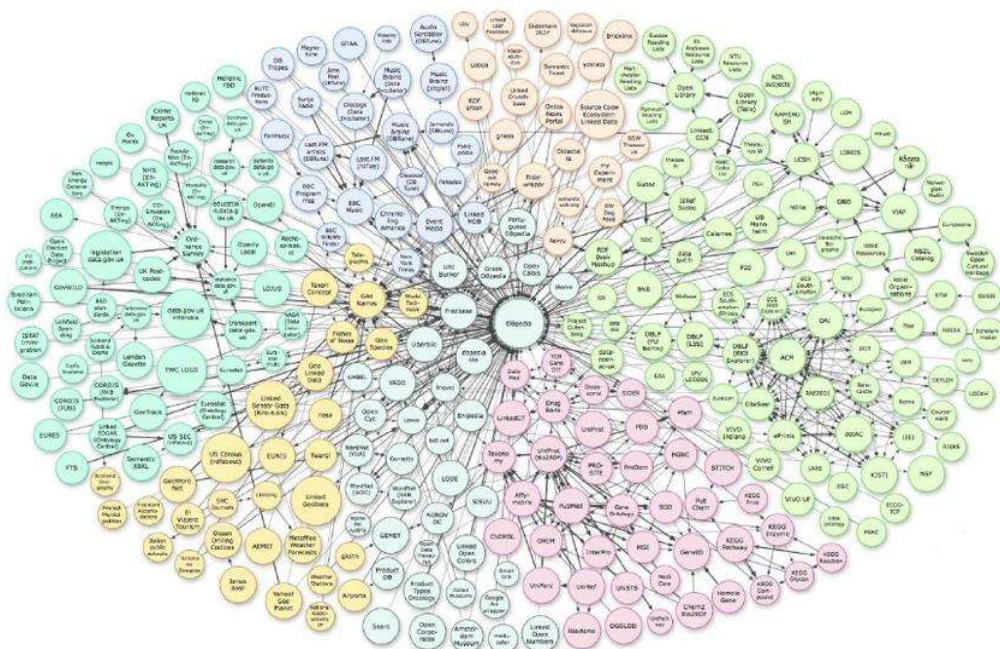
1. Матричну репрезентацију коришћењем матрица суседности
2. Уланчану репрезентацију коришћењем листа суседности
3. Секвенцијалну репрезентацију коришћењем линеаризованих листа суседности

Више информација о наведеним меморијским репрезентацијама графа се може пронаћи у материјалима са предавања и вежби, као и у књизи проф. Мила Томашевића „Алгоритми и структуре података“.

### Лексичка база знања

Моделовање људског знања у циљу разумевања текста од стране рачунара предмет је интензивног истраживања. Један од начина да се то знање моделује јесте усмереним тежинским графом, где се речи представљају као чворови, а семантичке релације одређених речи представљају као гране између њих. Тежина придружена грани моделује семантичку сличност тако да веће тежине грана представљају већу семантичку сличност. Тежине моделују реалним бројем у опсегу од 0 до 1.

Семантичке сличности могу представљати синониме, хипониме, мерониме, холониме, итд. Најсличнија реч одређеној речи је она која има највеће растојање од посматране речи. Уколико је реч А слична речи В са коефицијентом 0.19, то значи да постоји грана између А и В тежине 0.19. Ако посматрамо повезаност речи А и С преко речи В, коефицијенти сличности на том путу се множе.



Слика 1. Пример једне лексичке базе знања представљене графом

## Формирање и манипулација графом [50 поена]

Написати програм на програмском језику C++ који илуструје рад са усмерним тежинским графовима. Програм треба да омогући следеће операције над графом који представља једну лексичку базу знања:

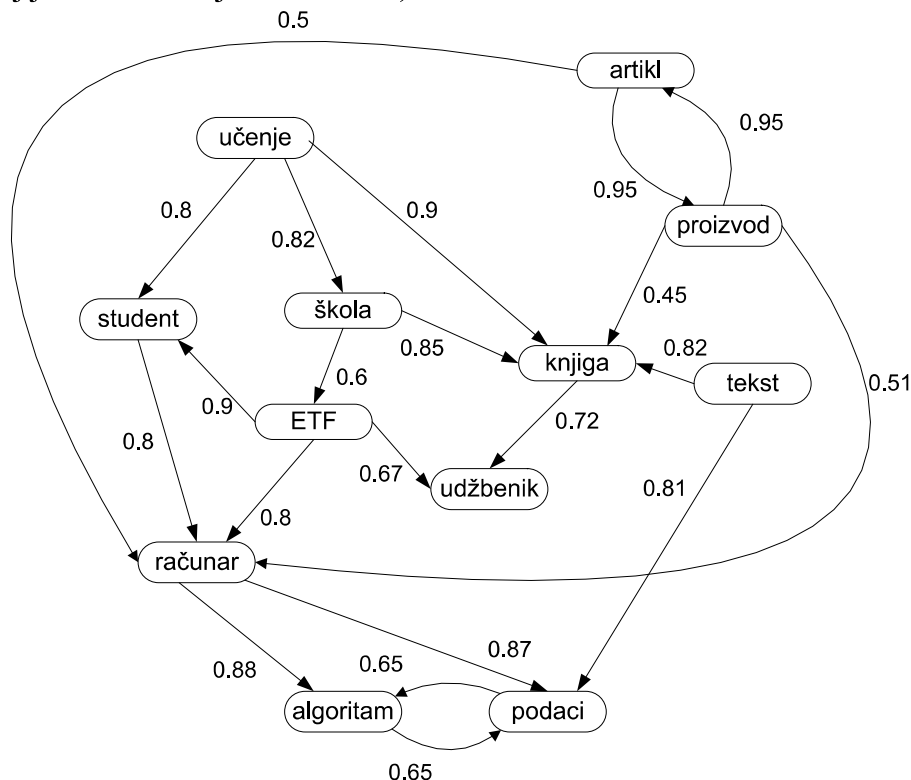
- [15 поена] Учитавање графа из текстуалног фајла задатог формата
- [10 поена] Додавање чвора у граф и уклањање чвора из графа
- [10 поена] Додавање и уклањање гране између два чвора у графу
- [5 поена] Испис репрезентације графа
- [10 поена] Брисање графа из меморије

Формат текстуалног фајла из којег се читају подаци је следећи: у првом реду се налази број чворова графа ( $n$ ), у другом реду број грана ( $e$ ), у трећем реду се налази  $n$  речи које су идентификатори чворова, а сваки од следећих  $e$  редова је облика *реч1 реч2 сличносћ*. Сматрати да реч не садржи више од 256 карактера, као и да су редови сортирани лексикографски. Два примера графа са слика 1 и 2 дата су у прилогу.

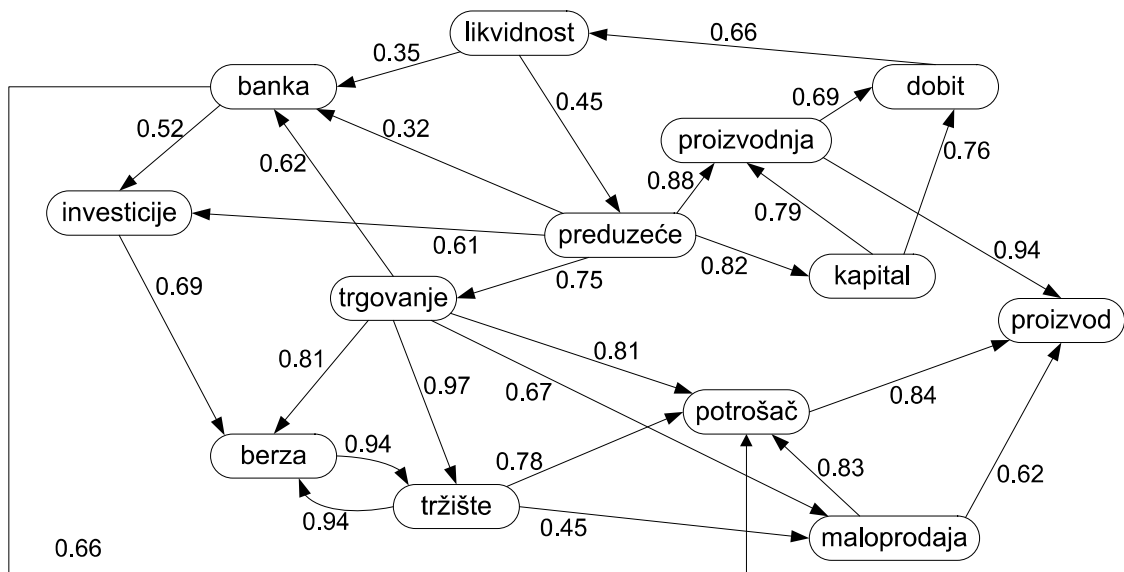
## Интеракција са базом знања [50 поена]

У циљу налажења повезаности одређених речи и анализе базе знања, потребно је омогућити и следеће операције над базом знања:

- [20 поена] Налажење  $k$  (уноси корисник) семантички најсличнијих речи задатој речи, у редоследу од најсличније ка најмање сличној (може их бити и мање од  $k$ , у зависности од повезаности чвора)
- [10 поена] Испис свих речи на најкраћем путу од задате једне речи до друге задате речи, у формату *реч1 -(сличносћ)->реч2 -(сличносћ)->реч3 ...* итд.
- [20 поена] Налажење свих речи које су јако повезане са задатом речју (припадају истој јако повезаној компоненти)



Слика 1. Граф који представља део лексичке базе општег знања



Слика 2. Граф који представља део лексичке базе за финансијски домен