

Универзитет у Новом Саду, Факултет техничких наука

ОАС Софтверско инжењерство и информационе технологије

Организација података

Организације датотека и методе приступа

Садржај

- **Основни појмови**
- Организовање података
- Организације датотека
- Процедуре над датотекама
- Методе приступа
- Ресурси

Основни појмови

- Датотека са строго дефинисаном структуром – нивои (Mogin, 2008)
 - логичка структура над скупом обележја
 - тип слога
 - тип слога је именовани низ обележја где је сваком обележју придружен домен
 - домен је тип податка
 - за тип слога се дефинише кључ
 - логичка структура над скупом података
 - скуп појава типа слога снабдевен релацијом строгог поретка
 - релација по којој су појаве уређене према вредности кључа
 - физичка структура података
 - подаци датотеке, информација о структури типа слога и евентуално информација о везама између слогова смештени на меморијски медијум

Основни појмови

- Датотека – концепти
 - поље
 - слог
 - блок

Основни појмови

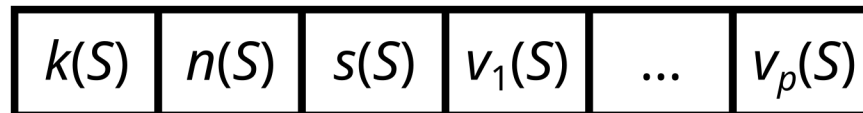
- Поље (Mogin, 2008)
 - поље је најмања организациона јединица података датотеке, која садржи смисао
 - поља садрже индивидуалне податке, конкретизације обележја и носе информацију о структури типа слога

ОСНОВНИ ПОЈМОВИ

- Слог (Mogin, 2008)
 - слог је појава типа слога уписана у датотеку
 - појава типа слога је у структурном погледу низ поља
 - у физичкој структури датотеке, слог може имати додатна поља у односу на одговарајућу појаву типа слога
 - додатна поља за статус слога и показиваче

ОСНОВНИ ПОЈМОВИ

- Слог ^(Mogin, 2008)
 - генеричка структура слога S
 - $k(S)$ – вредност кључа
 - $n(S)$ – конкретизација некључних обележја слога
 - $s(S)$ – статус слога
 - обично само индикатор актуелности слога
 - $v_i(S)$ – показивачи
 - $1 \leq i \leq p$



Основни појмови

- Слог – показивачи (Mogin, 2008)
 - показивач садржи информацију о адреси локације повезаног слога
 - један слог може садржати више показивача
 - показивачи се у датотеци могу појавити и изван слогова

Основни појмови

- Слог – показивачи – врсте адреса (Mogin, 2008)
 - **апсолутна адреса**
 - машинска адреса
 - пример за диск – (u, c, t, s)
 - ознака уређаја (u), редни број цилиндра (c), редни број стазе на цилиндру (t), редни број сектора на стази (s)
 - **релативна адреса**
 - редни број локације дела меморијског простора датотеке
 - $1 \leq rel_adr \leq Q$, Q је број локација у меморијском простору датотеке
 - интерно бива трансформисана у апсолутну адресу
 - **симболичка адреса**
 - вредност кључа повезаног слога
 - интерно бива трансформисана у релативну адресу

Основни појмови

- Блок (Mogin, 2008)
 - блок је организациона јединица података која представља групу слогова
 - блок представља јединицу размене података између диска и оперативне меморије
 - делови блока
 - један или више слогова
 - заглавље, по потреби
 - заглавље садржи податке о блоку
 - нпр. број актуелних слогова у блоку или број слободних локација у блоку
 - фактор блокирања f
 - број слогова за један блок ($f \geq 1$), константан на нивоу датотеке

ОСНОВНИ ПОЈМОВИ

- Блок (Mogin, 2008)
 - мотивација за увођење блокова
 - побољшање искоришћења меморијског простора
 - обично је капацитет сектора на диску већи од величине једног слога
 - повећање ефективне брзине размене података између оперативне меморије и екстерног меморијског уређаја
 - време размене података очекивано мање од средњег времена приступа
 - повећање ефикасности обраде података кроз смањење броја неопходних улазно-излазних операција
 - у случају да се слогови који су потребни у узастопним корацима обраде налазе у истом блоку

Основни појмови

- Датотека (Mogin, 2008)
 - један једноставан облик датотеке
 - датотека као низ блокова
 - блок као низ слогова
 - слог као низ поља
 - могућност увођења заглавља
 - блокови датотеке могу бити проширени својим заглављима
 - датотека може бити проширена заглављем

Основни појмови

- Датотека – заглавље (Mogin, 2008)
 - заглавље у облику специјалног слога смештеног на почетак датотеке
 - специјални слог садржи податке о структури и стању датотеке
 - нпр. број слогова у датотеци или показивач на слог од посебног значаја у датотеци
 - не мора свака датотека имати заглавље

ОСНОВНИ ПОЈМОВИ

- Датотека – ознака краја (Mogin, 2008)
 - чести начини за означавање краја датотеке
 - специјална ознака за крај уписана у прву слободну локацију датотеке
 - специјална ознака за крај уписана у показивач слога који је логички последњи
 - путем посебне евиденције о броју заузетих локација
 - означавање краја датотеке се не мора изводити у свакој датотеци

ОСНОВНИ ПОЈМОВИ

- Датотека – прорачун потребног простора (Mogin, 2008)

- K_d – капацитет меморијског простора потребан за смештај датотеке ако су слогови константног броја поља и константне дужине

$$K_d = \left\lceil \frac{x}{f} \right\rceil K_b + W_d$$

x – број слогова

$$x = N_o + N_s$$

f – фактор блокирања

$$f = \left\lceil \frac{K_b - K_z}{K_l} \right\rceil$$

K_b – капацитет физичког блока

W_d – капацитет потребан за податке које оперативни систем чува о датотеци (системска табела датотеке)

N_o – број обичних слогова

N_s – број специјалних слогова
 $N_s \in \{0, 1, 2\}$, зависно од тога да ли потребан слог за заглавље датотеке, слог за ознаку краја датотеке, оба или ниједан

K_z – капацитет локације за смештај заглавља блока

K_l – капацитет локације за смештај слога

ОСНОВНИ ПОЈМОВИ

- Датотека као резултат организовања података (Mogin, 2008)
 - физичка структура података, која одговара логичкој структури над скупом појава једног типа слога, снабдевена заглављем и потенцијално проширена помоћним структурама података

Садржај

- Основни појмови
- **Организовање података**
- Организације датотека
- Процедуре над датотекама
- Методе приступа
- Ресурси

Организовање података

- Организовање података је повезивање података у целину у циљу обезбеђења услова за њихову ефикасну обраду (Mogin, 2008)
 - организација података је назив за резултат организовања података
 - резултат организовања података је физичка структура података
 - резултат може бити скуп датотека
 - резултат може бити база података

Организовање података

- Организовање података – активности (Mogin, 2008)
 - активности у очекиваном редоследу извршења (није свака активност увек обавезна)
 - дефинисање скупа обележја
 - пројектовање логичке структуре над скупом обележја
 - избор поступка за доделу адреса локација слоговима
 - избор поступка за меморисање логичких веза између слогова
 - дефинисање помоћних структура података
 - прорачун и резервисање простора на медијуму меморијског уређаја
 - смештање слогова и њихових веза на меморијски медијум

Организовање података

- Организовање података – активности (Mogin, 2008)
 - **дефинисање скупа обележја**
 - изводи се на основу циљева/задатака за информациони систем
 - **пројектовање логичке структуре над скупом обележја**
 - дефинисање типова слогова
 - дефинисање релација између појава слогова различитог типа
 - одређивање кључева за сваки тип слога

Организовање података

- Организовање података – активности (Mogin, 2008)
 - **избор поступка за доделу адреса локација слоговима**
 - основни начини доделе локација слоговима
 - уписивање новог слога на крај датотеке
 - уписивање новог слога у прву слободну локацију
 - слободне локације су спрегнуте у ланац
 - уписивање новог слога у локацију чија релативна адреса представља функцију вредности кључа

Организовање података

- Организовање података – активности (Mogin, 2008)
 - **избор поступака за меморисање логичких веза између слогова**
 - основни начини за меморисање логичких веза између слогова
 - путем физичког позиционирања
 - логички суседни слогови бивају смештани у физички суседне локације
 - путем показивача
 - показивач показује на локацију непосредног логичког следбеника
 - место показивача
 - у формату слога
 - у помоћној структури података (нпр. у стаблу тражења)
 - не мора се увек водити рачуна о логичким везама између слогова

Организовање података

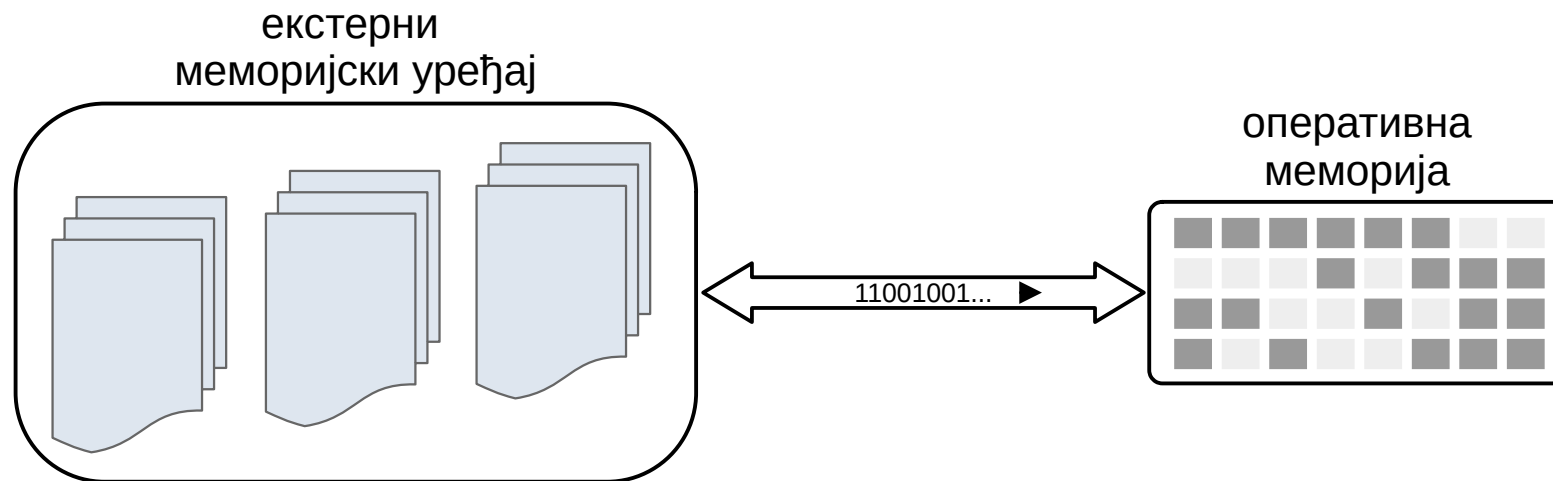
- Организовање података – активности (Mogin, 2008)
 - **дефинисање помоћних структура података**
 - специјалне помоћне структуре података не морају увек постојати
 - помоћне структуре података бивају увођене ради повећања ефикасности у употреби података, али подразумевају и увођење редувантности
 - **прорачун и резервисање простора на медијуму меморијског уређаја**
 - може бити обављено и динамички од стране оперативног система
 - **смештање слогова и њихових веза на меморијски медијум**

Садржај

- Основни појмови
- Организовање података
- **Организације датотека**
- Процедуре над датотекама
- Методе приступа
- Ресурси

Организације датотека

- Основни систем рада над датотекама (Mogin, 2008)
 - датотеке се обрађују у оперативној меморији (ОМ)
 - датотеке се смештају на екстерни меморијски уређај (ЕМУ)
 - обезбеђење перзистенције података
 - нижа цена меморисања података



Организације датотека

- Теорија о организацијама датотека (Mogin, 2008)
 - разумевање рада над структурама података у ОМ није довољно
 - посебна теорија је неопходна
 - због потребе преноса података између ЕМУ и ОМ
 - због разлика у времену приступа подацима између ЕМУ и ОМ
 - потребно узети у обзир специфичности саме врсте ЕМУ
 - због немогућности смештање целокупне датотеке у ОМ у појединим случајевима

Организације датотека

- Датотека као резултат организовања података (Mogin, 2008)
 - физичка структура података, која одговара логичкој структури над скупом појава једног типа слога, снабдевена заглављем и потенцијално проширена помоћним структурама података

Организације датотека

- Врсте организација датотека (Mogin, 2008)
 - постоје разне врсте организација датотека
 - разврставање може бити сагледано на основу више фактора
 - који је примењени поступак за доделу адреса слоговима
 - који је примењени поступак за меморисање логичких веза између слогова
 - да ли постоје помоћне структуре података и, ако постоје, каквих су карактеристика

Организације датотека

- Врсте организација датотека (Mogin, 2008)
 - примери
 - **серијска**
 - **секвенцијална**
 - **спрегнута**
 - **расута**
 - **индексна**
 - **с више кључева**

Организације датотека

- Врсте организација датотека (Mogin, 2008)
 - две групе врста организација датотека
 - **основне**
 - примери
 - серијска, секвенцијална, спрегнута, расута
 - **напредне**, које настају извођењем из основних
 - обухватају помоћне структуре података
 - примери
 - индексна, с више кључева

Организације датотека

- Врсте организација датотека – основне (Mogin, 2008)
 - **серијска**
 - слогови бивају распоређивани један за другим у редоследу у којем настају (хронолошки)
 - додела локација слоговима
 - уписивање новог слога на крај датотеке
 - меморисање логичких веза
 - /
 - помоћне структуре података
 - /



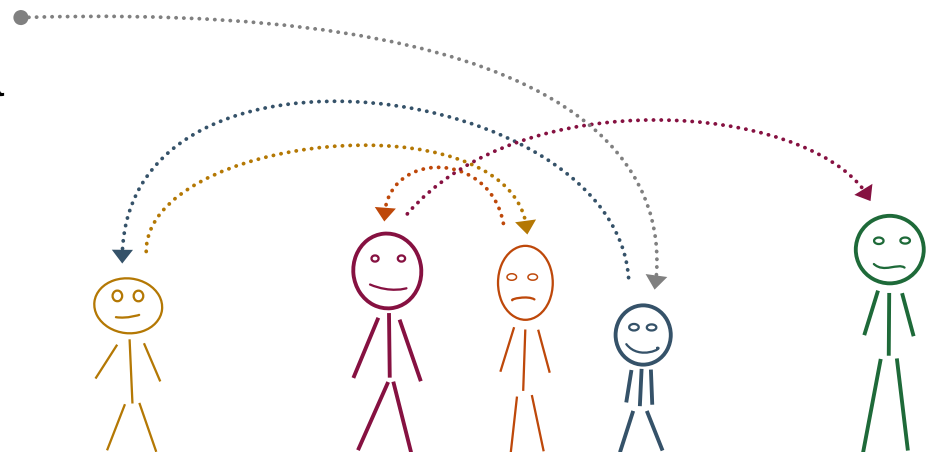
Организације датотека

- Врсте организација датотека – основне (Mogin, 2008)
 - **секвенцијална**
 - слогови бивају распоређивани према логичком поретку
 - обично у растућем поретку у односу на вредност кључа
 - додела локација слоговима
 - уписивање новог слога на крај датотеке (појавом нових слогова искоришћени простор се постепено проширује)
 - меморисање логичких веза
 - помоћу физичког позиционирања
 - помоћне структуре података
 - /



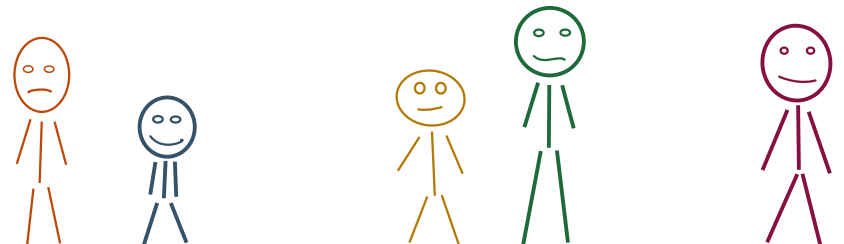
Организације датотека

- Врсте организација датотека – основне (Mogin, 2008)
 - **спрегнута**
 - слогови су спрегнути путем показивача и слободне локације су спрегнуте путем показивача
 - додела локација слоговима
 - уписивање новог слога у прву слободну локацију
 - слободне локације су спрегнуте у ланац
 - меморисање логичких веза
 - путем показивача
 - помоћне структуре података
 - /



Организације датотека

- Врсте организација датотека – основне (Mogin, 2008)
 - **расута**
 - слогови бивају смештани на основу резултата трансформације вредности кључа
 - додела локација слоговима
 - уписивање новог слога у локацију чија релативна адреса представља функцију вредности кључа
 - меморисање логичких веза
 - /
 - помоћне структуре података
 - присуство помоћних структура података зависно од конкретне подврсте расуте организације



Организације датотека

- Врсте организација датотека – основне (Mogin, 2008)
 - **расута**
 - врсте расуте организације
 - директна
 - релативна
 - статичка расута
 - динамичка расута

Организације датотека

- Врсте организација датотека – основне – преглед (Mogin, 2008)

ВРСТА ОРГАНИЗАЦИЈЕ	ПАРАМЕТАР ОРГАНИЗАЦИЈЕ		
	додела локација слоговима	меморисање логичких веза	помоћне структуре података
серијска	на крај датотеке	/	/
секвенцијална	на крај датотеке	физичким позиционирањем	/
спрегнута	у прву слободну локацију из посебног ланца	путем показивача	/
расута	на основу вредности трансформације вредности кључа	/	зависно од подврсте организације

Организације датотека

- Врсте организација датотека – напредне (Mogin, 2008)
 - **индексна**
 - подаци распоређени у зонама
 - две главне зоне
 - зона података (примарна зона)
 - обухвата слоге
 - обично серијски или секвенцијално организована
 - зона индекса
 - обухвата стабло за потребе брзог приступа слоговима
 - трећа зона може постојати у неким подврстама
 - зона прекорачења
 - обухвата слоге прекорачиоце
 - помоћне структуре података
 - стабло у зони индекса

Организације датотека

- Врсте организација датотека – напредне (Mogin, 2008)
 - **индексна**
 - врсте индексне организације
 - статичка индексна (индекс-секвенцијална)
 - динамичка индексна (с *B*-стаблом)

Организације датотека

- Врсте организација датотека – напредне (Mogin, 2008)
 - **с више кључева**
 - постоји више дефинисаних кључева
 - примарни кључ
 - алтернативни кључеви
 - као алтернативни кључеви појављују се секундарни кључеви
 - секундарни кључ је подскуп скупа обележја типа слога такав да није суперкључ типа слога
 - за секундарне кључеве бивају формиране помоћне структуре података које служе томе да претраживање по секундарним кључевима буде што ефикасније

Организације датотека

- Врсте организација датотека – напредне (Mogin, 2008)
 - **с више кључева**
 - у основи индексна организација с *B*-стаблом
 - подаци распоређени у зонама
 - зона података (примарна зона)
 - зоне индекса
 - број зона индекса обично одговара броју дефинисаних кључева
 - помоћне структуре података
 - стабла у зонама индекса

Организације датотека

- Врсте организација датотека – напредне (Mogin, 2008)
 - **с више кључева**
 - врсте организације с више кључева
 - мултилисна (мултилиста)
 - инвертована

Садржај

- Основни појмови
- Организовање података
- Организације датотека
- **Процедуре над датотекама**
- Методе приступа
- Ресурси

Процедуре над датотекама

- Процедуре за рад над датотекама (Mogin, 2008)
 - формирање
 - приступање
 - тражење
 - претраживање
 - обрада
 - ажурирање
 - реорганизовање

Процедуре над датотекама

- Процедура – формирање (Mogin, 2008)
 - смештање слогова на меморијски медијум у складу с одабраним поступцима за доделу адреса локација слоговима и за меморисање логичких веза између слогова
 - два начина формирања
 - кроз упис нових слогова при ажурирању
 - код серијске, динамичке расуте и индексне с *B*-стаблом
 - кроз посебан поступак формирања
 - код секвенцијалне, спрегнуте, статичке расуте и статичке индексне

Процедуре над датотекама

- Процедура – приступање (Mogin, 2008)
 - приступање локацији меморијског простора додељеног датотеци
 - начини приступа
 - секвенцијални
 - у наредном кораку рада приступа се локацији која је физички суседна у односу на текућу локацију
 - директни
 - у наредном кораку рада приступа се произвољној локацији
 - не узима се у обзир текућа локација
 - динамички
 - комбинација секвенцијалног и директног приступа
 - по приступу произвољној локацији сукцесивно се приступа физички суседним локацијама

Процедуре над датотекама

- Процедура – тражење (Mogin, 2008)
 - проналажење актуелног слога по задатој вредности кључа
 - аргумент тражења – задата вредност кључа
 - исход тражења
 - успешно
 - међу актуелним слоговима постоји тражени слог
 - неуспешно
 - међу актуелним слоговима не постоји тражени слог

Процедуре над датотекама

- Процедура – тражење (Mogin, 2008)
 - конкретни резултати тражења
 - индикатор успешности тражења
 - једини обавезан
 - тражени слог
 - релативна адреса локације на којој је тражење стало

Процедуре над датотекама

- Процедура – тражење (Mogin, 2008)
 - основне методе тражења
 - линеарно тражење
 - тражење праћењем показивача
 - на основу трансформације аргумента тражења у адресу
 - врсте тражења
 - тражење логички наредног слога
 - ослања се на праћење логичких веза између слогова
 - узима се у обзир где се претходно тражење зауставило, ако је оно уопште постојало
 - не мора подразумевати тражење логички непосредно наредног слога
 - тражење случајно одабраног слога

Процедуре над датотекама

- Процедура – претраживање (Mogin, 2008)
 - проналажење слогова по вредностима секундарних кључева
 - аргумент претраживања – задате вредности секундарних кључева
 - специфични облици упита за претраживање
 - конјунктивни облик
 - потребно да све задате вредности секундарних кључева буду адекватно заступљене у слогу
 - дисјунктивни облик
 - довољно да само једна задата вредност секундарног кључа буде адекватно заступљена у слогу
 - конкретни резултат претраживања – подскуп скупа слогова

Процедуре над датотекама

- Процедура – претраживање (Mogin, 2008)
 - претраживање би могло бити уопштено да дозволи проналажање по вредностима било којих обележја
 - питање перформанси претраживања
 - тражење представља посебан случај претраживања

Процедуре над датотекама

- Процедура – обрада (Mogin, 2008)
 - извршавање систематизованог низа операција, чији је циљ трансформација података датотека
 - у обради може учествовати произвољно много датотека

Процедуре над датотекама

- Процедура – обрада (Mogin, 2008)
 - улоге датотека у обради
 - према радњама читања и писања
 - врсте улога
 - улазна датотека
 - датотека из које се чита
 - излазна датотека
 - датотека у коју се пише
 - иста датотека може имати обе улоге (бити и улазна и излазна)

Процедуре над датотекама

- Процедура – обрада (Mogin, 2008)
 - улоге датотека у обради
 - према радњи тражења
 - врсте улога
 - водећа датотека
 - датотека која се користи за генерисање вредности кључа које служе као аргументи тражења над обрађиваном датотеком
 - обрађивана датотека
 - датотека над којом се изводи тражење по генерисаним вредностима кључа
 - иста датотека може имати обе улоге (бити и водећа и обрађивана)

Процедуре над датотекама

- Процедура – обрада (Mogin, 2008)
 - основни режими обраде
 - редоследна обрада
 - када се у наредном кораку обраде захтева приступ слогу у физички суседној локацији или проналажење логички наредног слога у односу на текући слог
 - директна обрада
 - када се у наредном кораку обраде захтева тражење случајно одабраног слога

Процедуре над датотекама

- Процедура – ажурирање ^(Mogin, 2008)
 - довођење података датотеке у сагласност са стварним стањем
 - активности ажурирања
 - упис новог слога
 - заједно с додавањем евентуалних веза према другим слоговима
 - модификација постојећег слога
 - модификација вредности оних обележја која нису део кључа
 - брисање постојећег слога
 - заједно с уклањањем евентуалних веза према другим слоговима
 - активности ажурирања претходи тражење
 - упис подразумева неуспешно тражење
 - модификација и брисање подразумевају успешно тражење

Процедуре над датотекама

- Процедура – реорганизовање (Mogin, 2008)
 - поновно формирање које се повремено изводи услед деградације перформанси обраде изазване појавом одређених неправилности у физичкој структури података
 - примери неправилности
 - присуство неактуелних слога (логички обрисаних)
 - извесни слогови се не налазе у узастопним локацијама, иако би то требало, већ су разбацани по простору датотеке

Процедуре над датотекама

- Перформансе обраде датотеке (Mogin, 2008)



– карактеристике идеалне организације датотеке

- минимално заузеће меморијског простора
- максимално брзо тражење случајно одабраног слога
- максимално брзо тражење логички наредног слога
- максимално брзо ажурирање
- реорганизовање непотребно

Процедуре над датотекама

- Перформансе обраде датотеке (Mogin, 2008)
 - идеална организација датотеке у пракси не постоји
 - постојеће организације датотека задовољавају само неке од карактеристика идеалне организације
 - појединачне организације датотека имају своје предности и мане
 - избор организације датотека за примену зависи од потреба и приоритета у конкретном случају
 - врсте организација датотека могу бити међусобно поређене по томе какве перформансе постижу у одређеним ситуацијама



Процедуре над датотекама

- Мере за оцену перформанси обраде датотеке (Mogin, 2008)
 - квантитативне мере за оцену перформанси
 - четири главне компоненте мере
 - активност која бива анализирана
 - тражење случајно одабраног слога, тражење логички наредног слога, ажурирање (упис, модификација, брисање)
 - исход активности
 - успешно, неуспешно
 - основни показатељ
 - број приступа диску, број поређења вредности у централној јединици (поређења аргумента тражења и вредности кључа)
 - варијанта показатеља
 - апсолутни, средњи

Процедуре над датотекама

- Мере за оцену перформанси обраде датотеке
 - квантитативне мере за оцену перформанси – пример

КОМПОНЕНТЕ МЕРЕ

ОСНОВНИ ПОКАЗАТЕЉ

- број приступа
- број поређења аргумента тражења и вредности кључа

АКТИВНОСТ КОЈА БИВА АНАЛИЗИРАНА

- тражење случајно одабраног слога
- тражење логички наредног слога
- ажурирање
 - упис слога
 - модификација слога
 - брисање слога

ВАРИЈАНТА ПОКАЗАТЕЉА

- апсолутни
- средњи

ИСХОД АКТИВНОСТИ

- успешно
- неуспешно

ПРИМЕР МЕРЕ

средњи број приступа за успешан упис слога

Садржај

- Основни појмови
- Организовање података
- Организације датотека
- Процедуре над датотекама
- **Методе приступа**
- Ресурси

Методе приступа

- Метода приступа ^(Mogin, 2008)
 - скуп рутина које се старају о изградњи и одржавању структуре датотеке и пружају комплетне услуге уписа и читања података из датотеке, строго водећи рачуна о логичкој структури слогова
 - рутине могу бити у разним окружењима
 - оперативни систем
 - систем за управљање базама података
 - независни кориснички пакет програма
 - припремљен за коришћење у неком програмском језику
 - савремена програмска окружења углавном нуде само базичне методе приступа
 - у случају додатних потреба, обично се ради имплементација одговарајућих метода приступа

Методе приступа

- Метода приступа ^(Mogin, 2008)
 - за различите врсте организација датотека најчешће су потребне посебне методе приступа
 - метода приступа намењена за одређену врсту организације ставља на располагање корисничким програмима услуге које омогућавају употребу датотека управо такве врсте организације
 - метода приступа нуди инструкције високог нивоа, које могу бити употребљаване у корисничком програму

Методе приступа

- Метода приступа ^(Mogin, 2008)
 - за различите врсте организација датотека најчешће су потребне посебне методе приступа
 - метода приступа сакрива од корисничког програма комплексност која се тиче размене података с датотеком и структуре датотеке
 - конверзија типова података
 - управљање баферима
 - блокирање и растављање блокова на слоге
 - иницирање улазно-излазних операција
 - проналажење на диску блока који садржи тражени слог
 - пренос траженог слога у радну зону корисничког програма

Садржај

- Основни појмови
- Организовање података
- Организације датотека
- Процедуре над датотекама
- Методе приступа
- **Ресурси**

Ресурси

- Извори и литература
 - Pavle Mogin. Strukture podataka i organizacija datoteka. 3. izdanje. Računarski fakultet (Beograd, Srbija), CET (Beograd, Srbija). 2008.
 - Glava 4. Osnovne strukture podataka datoteka
 - Glava 5. Fizičke strukture datoteka i postupci razmene podataka
 - Glava 6. Organizacija, traženje i obrada datoteke