

## ذخیره و بازیابی داده ها

جستجوی دودویی، مرتب سازی بر اساس کلید، شاخص گذاری

احمد خادم زاده  
[khademzaeh@mail.iust.ac.ir](mailto:khademzaeh@mail.iust.ac.ir)  
 دانشگاه آزاد اسلامی واحد مشهد - بهار ۱۳۸۳

## جستجو بر اساس کلید

- فرض کنید که در یک فایل که از رکوردهای با طول ثابت استفاده می کند، نیاز به جستجوی بر اساس کلید را داریم.
- اگر شماره نسبی رکورد مورد نظر (RRN) را داشته باشیم، می توانیم با استفاده از Seek کردن، مستقیماً سراغ رکورد مورد نظر برویم.
- اما در عمل، RRN را نداریم و باید رکورد مورد نظر (که کلید آن داده شده است) را در فایل جستجو نماییم.

[khademzadeh@iust.ac.ir](mailto:khademzadeh@iust.ac.ir)

## جستجو بر اساس کلید

- اگر فایل با توجه به کلید مرتب نشده باشد، باید جستجو در فایل از ابتدا و تا زمان رسیدن به رکورد مورد نظر انجام گیرد.
- اگر فایل مورد نظر را با استفاده از فیلد کلید آن مرتب کنیم، می توانیم در فایل جستجوی دودویی انجام دهیم.

[khademzadeh@iust.ac.ir](mailto:khademzadeh@iust.ac.ir)

## مقایسه جستجوی دودویی و جستجوی سریال

- جستجوی دودویی:  $O(\log N)$
- جستجوی ترتیبی:  $O(n)$
- اگر اندازه فایل دو برابر شود، در جستجوی خطی (ترتیبی)، زمان لازم دو برابر خواهد شد، اما در جستجوی دودویی یک واحد به زمان لازم اضافه خواهد شد.

[khademzadeh@iust.ac.ir](mailto:khademzadeh@iust.ac.ir)

## مرتب سازی بر اساس کلید

- فرض کنید که فایلی نیاز به مرتب شدن دارد. اما این فایل به اندازه ای بزرگ است که در حافظه اصلی جای نمی گیرد.
- برای مرتب کردن این فایل، فقط کلیدها مورد نیاز می باشد.
- فرض کنیم که تمام کلیدها در حافظه اصلی جای می گیرند.

[khademzadeh@iust.ac.ir](mailto:khademzadeh@iust.ac.ir)

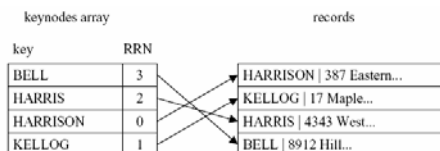
## مرتب سازی بر اساس کلید

- پس:
- کلیدها و شماره نسبی رکورد متناظر با هر کلید را به حافظه اصلی منتقل می کنیم.
- مرتب سازی داخلی کلیدها را انجام می دهیم.
- فایل را بصورت مرتب بازنویسی می کنیم.

[khademzadeh@iust.ac.ir](mailto:khademzadeh@iust.ac.ir)

## مرتب سازی بر اساس کلید

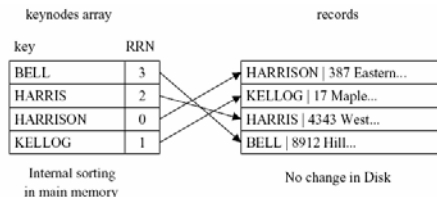
- فایل اولیه :



khademzadeh@iust.ac.ir

## مرتب سازی بر اساس کلید

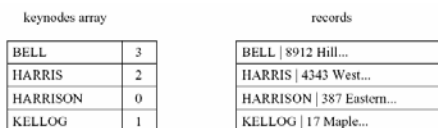
- مرتب سازی داخلی در حافظه اصلی (بدون تغییر در حافظه جانبی)



khademzadeh@iust.ac.ir

## مرتب سازی بر اساس کلید

- بازنویسی فایل بصورت مرتب شده :



khademzadeh@iust.ac.ir

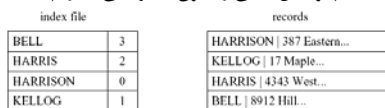
## هزینه انجام عمل مرتب سازی فایل

- هزینه لازم برای انجام این عمل (تعداد دسترسی به دیسک)
  - خواندن فایل بصورت ترتیبی
  - دسترسی به هر یک از رکوردها بصورت مستقیم (با Seek کردن)
  - نوشتن هر یک از رکوردها بصورت ترتیبی

khademzadeh@iust.ac.ir

## شاخص گذاری

- چرا فایل را مجدداً بازنویسی کنیم ؟
- از آرایه مرتب شده کلیدها به عنوان یک فایل شاخص (Index) استفاده کنیم. و فایل اصلی را بدون تغییر باقی گذاریم.



- به عمل فوق شاخص گذاری گفته می شود.

khademzadeh@iust.ac.ir

سوال ؟