

ذخیره و بازیابی اطلاعات

○ مهندس آزاده پورغفاری لاهیجی

- ساختارهای فایل
- مایکل فولک
- مترجم: علیرضا جباریه
- انتشارات جهان نو
- چاپ اول، ۱۳۷۹

اولین ویرایش کتاب «ساختارهای فایل» دربارهٔ ذخیره‌سازی ساختارهای فایل تا سال ۱۹۸۰ بود. ولی ویرایش دوم دربارهٔ انواع ذخیره‌سازی‌های مطرح شده تا به امروز (از دیسک‌های نوری گرفته تا DVD) می‌باشد. تاریخچه و سازماندهی ساختارهای فایل، عمده‌ترین مطلب این کتاب است ولی در متن کتاب به توسعهٔ سخت‌افزار و نرم‌افزار سیستمی نیز توجه شده است. در سی سال اخیر کامپیوترها از حالت یکپارچه خارج شده و به صورت کامپیوترهای مستقل (PC) می‌توانند عمل کنند. بنابراین نیاز است اطلاعات دریافتی، در مقصد، تجزیه و تحلیل شوند. امروزه برنامه‌نویسان ساختارهای فایلی را خودشان طراحی و ایجاد می‌کنند. این کتاب چگونگی طراحی و ایجاد ساختارهای فایلی کارآمد را نشان می‌دهد. برای آشنایی با طراحی ساختار فایل‌ها و انجام کارهای مرتبط با آن، نیاز به یادگیری یک زبان برنامه‌نویسی، سیستم عامل مناسب و ابزارهای مفهومی دارید. شش فصل اول این کتاب در ارتباط با ابزارهای اصلی برای طراحی ساختارهای فایلی ساده است. در متن کتاب مثال‌هایی به صورت کدهای برنامه‌نویسی آورده شده است و اگر کاربر UNIX باشید و چگونگی و نحوهٔ ساختار و سیستم فایل‌ها را مشاهده خواهید کرد. فصل اول به آشنایی با ساختارهای فایل می‌پردازد، آشنایی با موضوع طراحی اولیه که طراحی ساختار فایل را مشخص می‌کند. همچنین بر تاریخچه طراحی ساختار فایل مروری دارد، زیرا پی‌گیری توسعه ساختارهای فایل، اطلاعات بیشتری دربارهٔ طراحی ساختارهای فایل خودمان به ما می‌آموزد، آشنایی با نکات یادگیری

ساختار فایل و مجموعهٔ مفاهیم طراحی ساختار فایل، از هدف‌های رفتاری این فصل می‌باشند. در فصل دوم اساس عملیات پردازش فایل مورد بررسی قرار می‌گیرد. در این فصل می‌آموزید فرآیند پیوند یک فایل منطقی درون برنامه به یک فایل فیزیکی یا یک ابزار را شرح دهید روال‌های مورد استفاده برای ایجاد، باز کردن و بستن فایل‌ها را توصیف کنید، زیر روال‌های مورد استفاده برای خواندن و نوشتن فایل‌ها را شرح دهید، آشنایی با مفهوم Position (موقعیت) داخل یک فایل و زیرروال‌های جستجوی موقعیت‌های مختلف را شرح دهید. آشنایی با سازماندهی سیستم فایلی UNIX، دیدگاه UNIX نسبت به فایل را نمایش داده و عملیات فایلی UNIX و فرمان‌های براساس این دیدگاه را شرح دهید. در فصل سوم رسانهٔ ذخیره‌سازی ثانویه و نرم‌افزار سیستم شرح داده می‌شود. پس از مطالعه این فصل بایستی بتوانید درایورهای دیسک که شامل واحدهای اساسی و ارتباط آنها می‌باشد را تشریح کنید، عوامل مؤثر در زمان دسترسی دیسک و روش‌های برآورد زمان‌های دسترسی و فضای مورد نیاز و سرعت انتقال را شناسایی و تشریح کنید، تفاوت‌های اصلی بین رسانه‌ها و معیارهایی که می‌توانند برای هماهنگی مستقیم رسانه با یک برنامه کاربردی مورد استفاده قرار گیرند را شناسایی کنید. واژه‌های عمومی رویدادهایی که هنگام انتقال بین برنامه و رسانهٔ ذخیره‌سازی، رخ می‌دهند، را شرح دهید. مفاهیم و تکنیک‌های مدیریت بافر را معرفی کنید. چند مفهوم مقدماتی علی‌الخصوص مفاهیم نرم‌افزار سیستم در ارتباط با یونیکس را شرح دهید. فصل چهارم به مفاهیم اصلی ساختار فایل می‌پردازد، در این فصل آشنایی با مفاهیم ساختار فایل، فایل‌های جریان، محدودیت‌های فیلد و رکورد، فیلدها و رکوردهای با طول ثابت و طول متغیر، کلیدهای جستجو و شکل‌های استاندارد، جستجوی ترتیبی، دسترسی مستقیم، دسترسی به فایل و سازماندهی فایل، بررسی انواع دیگر ساختارهای فایل

با عباراتی مثل، مدل‌های داده‌های مجرد، فوق داده‌ها (دادگان) دسترسی به فایل شیء‌گرا، توسعه، بررسی موضوعات قابلیت حمل و استانداردسازی شرح داده می‌شود. فصل پنجم سازماندهی فایل‌ها به منظور افزایش کارایی نامیده شده است. این فصل مروری بر چند روش فشرده‌سازی داده‌ها دارد، نگاهی دارد به بررسی یک روش ساده مثل «فشرده‌گی منبع ذخیره‌سازی» به عنوان راهی ساده برای استفاده مجدد از فضای فایل، معرفی یک ریز برنامه برای حذف رکوردهایی با طول ثابت که باعث اتلاف فضای فایل می‌گردند و به صورت پویا می‌توانند مورد استفاده مجدد قرار گیرند، شرح استفاده از لیست‌های پیوندی و پشته‌ها برای مدیریت لیست موجود، بررسی چند روش برای حذف رکوردهای با طول متغیر، بررسی چند روش برای حذف رکوردهای با طول متغیر، آشنایی با مفهوم پیوستگی با اصطلاحات «تکه‌تکه شدن داخلی» و «تکه‌تکه شدن خارجی» تشریح راه‌کارهای جایگزینی مربوط به استفاده مجدد از فضای فایل‌های حاوی رکوردهای با طول متغیر، تشریح روش جستجوی دودویی، بررسی محدودیت‌های جستجوی دودویی، معرفی و ایجاد ریز برنامه‌ها «مرتب‌سازی کلیدی» برای مرتب‌سازی فایل‌های بزرگ و بررسی هزینه‌های آن و آشنایی با مفهوم «رکوردهای متصل».

دو فصل بعد، آشنایی با مهم‌ترین طراحی‌های ساختار فایل سطح بالا است که شامل درخت‌های B و B+ می‌باشد.

فصل ششم مروری دارد بر آشنایی با مفهوم شاخص‌گذاری که دارای کاربردهای گسترده‌ای در طراحی سیستم‌های فایلی است، آشنایی با کاربرد شاخص خطی ساده برای ارائه دسترسی سریع به رکوردها در یک ورودی ترتیبی و فایلی با رکورد طول متغیر، بررسی دلایل استفاده از شاخص‌ها برای نگهداری فایل، شرح کاربرد شاخص‌ها برای فراهم نمودن دسترسی به رکوردها توسط بیش از یک کلید، آشنایی با نظریه لیست معکوس و شرح عملیات

سیستم‌های mainframe که برای مرتب‌سازی فایل‌های بزرگ به کار می‌رفتند، دارای ۲ تا ۴ مگابایت حافظه اولیه بودند ولی حالا ۳۲ تا ۶۴ مگابایت روی ایستگاه‌های کاری، متداول است و کامپیوترهایی با چندین گیگا بایت حافظه RAM وجود دارند. هنگامی که RAM بیشتری موجود است، می‌توان مشکلات ساختارهای فایلی را به طور متفاوت بررسی کرد. به عنوان مثال، کتاب‌های ساختار فایل قبلی بیشتر درباره مرتب‌سازی فایل‌های بزرگ روی نوار مغناطیسی بحث می‌کردند. یکی از دلایل این امر، این است که اگر مقدار RAM کم باشد، مرتب‌سازی روی نوار مغناطیسی، پایدارتر از مرتب‌سازی روی دیسک است. اکنون که RAM ارزان‌تر و قابل دسترس‌تر است، مرتب‌سازی روی دیسک، نه تنها پایدار است، بلکه معمولاً ترجیح نیز داده می‌شود.

فصل هشتم به معرفی فایل‌های ترتیبی شاخص‌دار، توضیح عملیات روی مجموعه ترتیبی از بلاک‌هایی که رکوردها را به ترتیب براساس کلید نگه می‌دارند، نمایش چگونگی ایجاد مجموعه شاخص در مجموعه‌ای ترتیبی برای تولید یک ساختار فایل ترتیبی شاخص‌دار، آشنایی با کاربر درخت B برای نگهداری مجموعه شاخص و در نتیجه آشنایی با درختان B+ و درختان B+ پیشوندی ساده، چگونه مجموعه شاخص درخت B در یک درخت B+ ساده می‌تواند مرتبه متغیر داشته و تعداد متغیری جداکننده را نگهداری نماید و مقایسه مزایا (نقاط قوت) و معایب (نقاط ضعف) درختان B پیشوندی ساده، درختان B+ و درختان B می‌پردازد. کتاب در حقیقت، در دو بخش و مطابق با سرفصل درس‌های ذخیره و بازیابی ۱ و ۲ جهت استفاده دانشجویان گروه کامپیوتر تنظیم و نگارش شده است.



صفحه‌ای و درختان B مجساز و توضیح الگوریتم‌های متفاوت و اصلی درخت B مثل آنهايي که برای ایجاد B و درختان B با رکوردهای طول متغیر، مورد استفاده قرار می‌گیرند، از موضوعات این فصل می‌باشند.

توسعه طراحی نرم‌افزار در ده ساله اخیر، دلیل کافی برای ویرایش دوم کتاب است. ضمناً در این ویرایش درباره سخت‌افزار ذخیره‌سازی نیز شرح داده شده است. به عنوان مثال یکی از تغییرات آرمانی در پی‌کربندی کامپیوترها افزایش مقدار RAM موجود در انواع کامپیوترها بود. هنگامی که نوشتن اولین ویرایش این کتاب به پایان رسید، حداکثر RAM موجود در کامپیوترهای شخصی، ۶۴۰ کیلوبایت بود. اکنون برای ساده‌ترین برنامه کاربردی حداقل ۴ مگابایت حافظه مورد نیاز است. سال‌ها پیش

منطقی (Boolean) روی لیست‌ها، تحلیل این موضوع که کلید شاخصی به آدرسی در فایل داده محدود می‌شود و آشنایی و بررسی فایل‌های خود شاخص. فصل هفتم درختان B و سایر سازماندهی فایلی مبتنی بر درخت (شاخص‌بندی چند سطحی) نامیده شده است. توسعه درختان B برای حل مشکلات و مسایلی که این درختان برای آنها طراحی شده‌اند، مروری خلاصه بر سایر ساختارهای درختی که ممکن است در رسانه ذخیره‌سازی مورد استفاده قرار گیرند مثل درختان AVL صفحه‌بندی شده، ارائه مفهوم ویژگی‌های مهم که توسط درختان B فراهم می‌شود و نشان دادن این موضوع که چگونه این ویژگی‌ها علی‌الخصوص برای کاربردهای حافظه جانبی، مناسب هستند، شرح عملیات اساسی روی درختان B، آشنایی با فرهنگ