

۱۴۹۴۹۱۵
۹۶,۲,۱۱

ذخیره و بازاریابی اطلاعات

www.ketab.ir

مؤلفان:

الهام دانشمند و ژیلدا حسینی

حزیم دانش

سرشناسه : دانشمند، الهام، ۱۳۵۲ -
 عنوان و نام پدیدآور : ذخیره و بازیابی اطلاعات / مولفین: الهام دانشمند و ژيلا حسيني؛
 ویراستار مریم سادات حسینی
 مشخصات نشر : تهران: حریم دانش، ۱۳۹۶.
 مشخصات ظاهری : ۲۰۴ ص.: مصور، جدول.
 شابک : ۹۷۸-۹۶۴-۲۷۰۶-۳۹-۰۰
 وضعیت فهرست نویسی : فیفا
 موضوع : ساختار داده‌ها - ذخیره و بازیابی اطلاعات

موضوع: Data structures (Computer science)-Information storage and retrieval systems

رده‌بندی کنگره : QA76/۹/س۲د۲ ۱۳۹۶۰

رده‌بندی دیویی : ۰۰۵/۱۳

شماره کتابشناسی ملی : ۴۶۷۹۱۱۷۰

عنوان کتاب: ذخیره و بازیابی اطلاعات

مولفان: الهام دانشمند و ژيلا حسيني

ویراستار: مریم سادات حسینی

ناشر: حریم دانش

نوبت چاپ: اول ۱۳۹۶

قیمت: ۱۵۰,۰۰۰ ریال

تیراژ: ۱۰۰۰ جلد

شابک: ۹۷۸-۹۶۴-۲۷۰۶-۳۹-۰۰

کلیه حقوق چاپ و نشر برای انتشارات حریم دانش محفوظ می‌باشد.

مرکز پخش: میدان انقلاب، ضلع جنوب شرقی میدان انقلاب، پلاک ۱۱ جدید، کتابفروشی

گلریزان. تلفن: ۰۹۱۲۳۴۶۰۸۵۹ - ۶۶۹۷۶۲۳۰

برای خرید کتابهای انتشارات حریم دانش می‌توانید به سایت Harim-Danesh.ir نیز مراجعه کنید.

فهرست مطالب

۱۵	 مقدمه
۱۷	 پیشگفتار
۲۱	 فصل اول: سیر تکامل دستگاه‌های ذخیره سازی اطلاعات
۲۲	 سیر تکامل دستگاه‌های ورودی
۲۳	 کارت پانچ
۲۵	 نوار کاغذی
۲۶	 نوار مغناطیسی
۲۸	 دیسک مغناطیسی
۲۹	 دیسکهای نوری
۳۰	 کاراکتر چیست؟
۳۱	 اطلاعات و نحوه‌ی سنجش آنها
۳۳	 انواع وسایل متداول ذخیره‌سازی اطلاعات (Types of Storage Media)
۳۳	 دیسک سخت داخلی یا Internal Hard Disk
۳۵	 دیسک سخت خارجی (External Hard Disk)
۳۵	 سی دی (CD)
۳۶	 DVD (Digital Versatile Disk)
۳۷	 دیسکهای بلو-ری (Blu-Ray)
۳۷	 حافظه فلش (Flash Memory)
۳۸	 کارتهای حافظه (Memory Cards)
۳۸	 فلاپی دیسک (Floppy Disk)
۳۹	 دیسکهای فشرده (Zip Disks)

۳۹	نوار مغناطیسی یا کارتریج داده‌ها (Data Cartridges)
۴۰	Online درایوهای شبکه و ذخیره‌سازی
۴۰	عوامل موثر در کارایی و سرعت یک کامپیوتر
۴۰	سرعت CPU (CPU Speed)
۴۱	اندازه حافظه‌ی اصلی (Ram Size)
۴۳	سرعت دیسک سخت (Hard Disk Speed)
۴۴	فضای آزاد دیسک (Free Hard Disk Space)
۴۴	یک بارچ‌سازی (Defragmenting Files) دیسک سخت
۴۶	تمرین
۴۷	فصل دوم: رسانه‌های ذخیره‌سازی
۵۰	تعریف عام حافظه
۵۱	خصوصیات حافظه در معنای عام
۵۳	سلسله مراتب حافظه‌ها
۵۶	انواع حافظه‌های برون ماشینی از نظر تکنولوژی ساخت
۵۷	نوار مغناطیسی
۵۷	معرفی رسانه
۵۸	نحوه ذخیره‌سازی داده روی نوار
۵۸	بیت پاریتی چیست؟
۵۹	چگالی نوار
۶۰	گپ (فاصله)
۶۰	نحوه ذخیره‌سازی فایل روی نوار
۶۱	پارامترهای نوار

۶۲	 دیسک مغناطیسی
۶۲	 معرفی رسانه
۶۲	 رده بندی دیسکها
۶۳	 تقسیمات دیسک
۶۵	 پارامترهای دیسک
۶۷	 نحوه ضبط داده ها روی دیسک
۶۷	 دیسک رم
۶۸	 دیسک های نوری
۶۹	 دیسک های تغییر فاز
۷۰	 دیسکهای دای-پول
۷۰	 طبقه
۷۰	 معرفی رسانه
۷۱	 موارد استفاده
۷۲	 تمرین
۷۳	 فصل سوم: مفاهیم و تعاریف پایه
۷۶	 مفاهیم مقدماتی
۷۶	 فیلد
۷۶	 رکورد
۸۳	 رکورد ذخیره شده (در محیط ذخیره سازی)
۸۸	 کلید رکورد
۸۹	 فایل: تعریف کلی
۹۰	 فایل: مکانیسم فیزیکی ذخیره سازی

۹۰		بلاک بندی
۹۱		تکنیکهای تعیین محدوده رکورد در بلاک
۹۲		مزایای بلاک بندی
۹۲		معایب بلاک بندی
۹۴		چگالی لود اولیه
۹۴		مزایای معایب
۹۵		موضعی بودن رکورد ها (لوکالیتی)
۹۶		درجات لوکالیتی
۹۷		فایل در محیط فیزیکی
۹۸		نشست پیوسته
۹۸		نشست ناپیوسته
۹۹		مدیریت بلاکهای آزاد
۹۹		آشنایی با بافر و بافرینگ
۱۰۰		نحوه ایجاد بافرها
۱۰۰		بافر از نظر محل ایجاد
۱۰۱		انواع بافرینگ
۱۰۳		تکنیک های تولید نسخه پشتیبان
۱۰۴		تکنیک ۱: استفاده از نیمه دو دیسک
۱۰۴		تکنیک ۲: تولید دامپهای تدریجی
۱۰۴		تکنیک ۳: آینه سازی (نظیر سازی)
۱۰۵		تمرین

۱۰۷	فصل چهارم: ساختارهای مبنایی فایل
۱۱۰	ضوابط ارزیابی کارایی
۱۱۲	متوسط اندازه رکورد
۱۱۸	شرح اصول عملیات ششگانه
۱۱۸	واکنش رکورد دلخواه
۱۲۱	بازیابی رکورد بعدی
۱۲۳	بهنگام سازی از طریق درج
۱۲۴	بهنگام سازی از طریق تغییر محتوای رکورد
۱۲۵	خواندن تمام ایل
۱۲۵	سازماندهی مجدد
۱۲۷	شرح ساختارهای مبنایی
۱۲۷	فایل با ساختار «پایل» (برهم)
۱۲۷	معرفی ساختار
۱۲۸	موارد استفاده
۱۲۸	ارزیابی کارایی
۱۳۱	فایل با ساختار ترتیبی
۱۳۱	معرفی ساختار
۱۳۳	موارد استفاده
۱۳۴	انواع روشهای جستجو
۱۳۷	فصل پنجم: ساختار فایل شاخص‌دار
۱۳۸	مفاهیم مبنایی
۱۳۸	شاخص

۱۳۹	شاخص اصلی و شاخص (های) ثانوی
۱۳۹	لنگرگاه (نقطه اتکاء)
۱۴۰	گروه در شاخص غیرمترکم
۱۴۰	شاخص خوشه ساز
۱۴۱	شاخص چندسطحی
۱۴۱	معیار شاخص بندی
۱۴۱	ساختار رتیبی شاخصدار
۱۴۱	معرفی ساختار
۱۴۲	بررسی مشکل
۱۴۵	موارد استفاده ساختار
۱۴۵	فایل چندشاخصی
۱۴۵	معرفی ساختار
۱۴۵	اجزاء اصلی ساختار
۱۴۶	فایل وارون
۱۴۶	نشست رکوردها روی رسانه
۱۴۶	ضابطه انتخاب صفات خاصه شاخص
۱۴۷	ساختار شاخص
۱۴۷	درج رکورد
۱۴۸	حذف رکورد
۱۴۹	موارد استفاده
۱۴۹	بهبودهایی در پیاده سازی شاخص
۱۵۱	تمرین

۱۵۳	فصل ششم: ساختار فایل مستقیم.....
۱۵۴	فایل مستقیم مبنایی.....
۱۵۵	برخی ویژگیهای ساختار.....
۱۵۶	لود کردن فایل.....
۱۵۷	لزوم اعمال تابع مبدل.....
۱۵۷	توابع مبدل.....
۱۵۸	انواع تابع درهم ساز.....
۱۵۹	ضوابط انتخاب تابع درهم ساز.....
۱۶۰	مشکل تصادف.....
۱۶۰	شرایط بروز پدیده تصادف.....
۱۶۰	احتمال تصادف.....
۱۶۱	باکت بندی.....
۱۶۳	راه حلهای مشکل تصادف.....
۱۶۳	راه حل اول: ایجاد یک فایل جداگانه.....
۱۶۴	معایب فایل مستقیم مبنایی.....
۱۶۴	موارد استفاده فایل.....
۱۶۵	روشهای بهبود ساختار مستقیم مبنایی.....
۱۶۶	تمرین.....
۱۶۷	فصل هفتم: ساختار چندحلقه ای.....
۱۶۸	معرفی ساختار.....
۱۶۸	نمایش ساختار در طراحی.....
۱۶۸	ساختار یک نوع رکورد.....

۱۶۹		سرآیند و نقش آن
۱۷۰		ساختار حلقه از نظر پیاده سازی
۱۷۰		ملاحظات در نشست فایل روی دیسک
۱۷۰		روش سیستم فایل در پاسخگویی به پرس و جوها
۱۷۰		موارد استفاده ساختار
۱۷۱		ارزیابی کارایی ساختار
۱۷۱		متدهای اندازه رکورد
۱۷۱		زمان واکنشی یک رکورد
۱۷۲		بازیابی رکورد
۱۷۳		عمل درج
۱۷۴		عمل بهنگام سازی
۱۷۵		خواندن تمام فایل
۱۷۶		سازماندهی مجدد
۱۷۷		روشهای بهبود کارایی ساختار چند حلقه ای
۱۷۸		ایجاد حلقه با نشانه روی های دو راهه
۱۷۸		ایجاد حلقه با نشانه روی هایی از اعضاء به سرآیند
۱۷۸		ترکیب ساختار چند حلقه ای با ساختار(های) دیگر
۱۷۹		فصل هشتم: ساختارهای درختی
۱۸۰		فایل با ساختار درخت جستجوی دودویی
۱۸۲		عملیات در فایل
۱۸۲		واکنشی رکورد
۱۸۳		عمل درج

۱۸۴	حذف رکورد.....
۱۸۴	خواندن تمام فایل.....
۱۸۵	فایل با ساختار درخت جستجوی دودویی نخ کشی شده (TBST).....
۱۸۶	فایل با ساختار درخت صفحه بندی شده.....
۱۸۷	فایل با ساختار درخت متعادل (B-TREE).....
۱۸۷	معرفی ساختار.....
۱۸۸	عملیات در ساختار.....
۱۸۹	واکنشی مورد.....
۱۸۹	خواندن سریال تمام فایل.....
۱۸۹	عمل درج.....
۱۹۱	عمل حذف.....
۱۹۲	فایل درختی.....
۱۹۲	شرح ساختار.....
۱۹۳	زمان واکنشی رکورد.....
۱۹۵	متعادل کردن درخت.....
۱۹۵	فایل با ساختار درخت B^+
۱۹۷	فایل با ساختار درخت k-d.....
۱۹۸	فایل با ساختار ترای.....
۱۹۹	بهبود در ساختار.....

مقدمه

بعد از ابزارها، ماشین‌ها، کامپیوتر سومین دستگاه مفیدی است که به دست بشر اختراع شده است. ابزارها و ماشین‌ها مستقیماً توسط اشیاء فیزیکی کار می‌کنند در حالی که کامپیوتر با عاملی بسیار منطقی که همان اطلاعات است عمل می‌کند. کامپیوتر با تمام پیچیدگی‌های ساخت آن صرفاً یک ابزار است که توسط بشر وجود آمده و در مدت عمر کوتاه خود به طور شگفت‌انگیزی تکامل یافته است.

اولین کامپیوترها صرفاً جهت کارهای محاسباتی و تحقیقاتی ساخته شدند و استفاده از آنها بسیار سخت و انحصاری بود و فقط افراد متخصص و رفقای قادر به استفاده از آنها بودند. ولی با گذشت زمان و اضافه شدن کاربردهای جدید به کامپیوتر و خارج شدن آن از حالت تخصصی، استفاده از کامپیوتر همگانی شده و سعی می‌شود استفاده از نامیوس آسانتر گردد.

در حال حاضر از کامپیوتر برای انجام کارهای مختلفی استفاده می‌شود ولی یکی از کاربردهای مهم آن نگهداری اطلاعات می‌باشد. در تعاریف مختلف برای کامپیوتر در زمینه‌های مختلفی که اطلاعاتی را از محیط بیرون دریافت نماید و بر اساس دستوراتی که از قبل به آن داده شده پردازشهایی (عملیات محاسباتی و مقایسه‌ای) را بر روی آن انجام دهد و نتایج حاصله را در خود ذخیره نموده یا به ما نمایش دهد کامپیوتر نامیده می‌شود.

با توجه به این تعریف کامپیوتر حداقل از سه بخش ورودی، پردازشگر و خروجی تشکیل شده است و واحد پردازشگر شامل حافظه نیز می‌باشد. البته در یک تقسیم‌بندی دیگر کامپیوتر را دارای سه بخش: ورودی-خروجی، پردازشگر و حافظه می‌دانند. در این کتاب ما می‌خواهیم

نگاهی موشکافانه به بخش حافظه داشته باشیم و حافظه را از جهات مختلف بررسی کنیم.

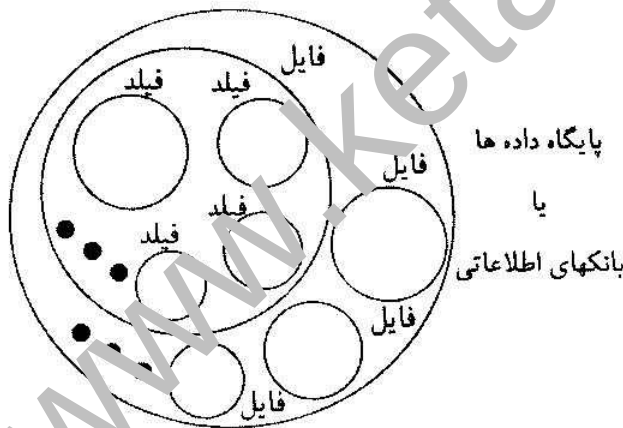
این بررسی‌ها در سه بخش انجام خواهد شد:

۱- شناخت انواع رسانه‌های ذخیره سازی

۲- تحلیل اصطلاحات رایج در ذخیره و بازیابی اطلاعات

۳- شناخت روش‌های پایه‌ای ذخیره سازی اطلاعات

مطالب این درس در خصوص ساختن و مدیریت فایل‌ها خواهد بود و در درس پایگاه داده‌ها رابطه بین فایل‌ها بررسی خواهد شد. اگر بخواهیم شکل نمادین زیر را در نظر بگیریم جایگاه درس ذخیره سازی تا مرحله ساخت فایل و مدیریت صحیح فیلدها خواهد بود که این مطالب پیش نیاز درس پایگاه داده می‌باشد.



در این درس به دنبال بررسی بهترین شکل طراحی ساختار رکورد و مزایا و معایب شیوه‌های مختلف ذخیره اطلاعات خواهیم بود.

پیشگفتار

قبل از شروع بحث، گذری خواهیم داشت به مباحث مبانی کامپیوتر و به طور مختصر پیدایش کامپیوتر و سیر تحول آن را مرور خواهیم کرد تا زمینه‌ساز درک بهتر سایر مطالب گردد. بعد از جنگ جهانی دوم به خاطر درگیری روز افزون بشر به کارهای اداری و تجاری با حجم زیاد و محاسبات پیچیده و حجم علمی، باعث بوجود آمدن جنبشی برای ساختن ماشین‌های سریع‌تر و قوی‌تر گردید. در حین انجام این فعالیتها و پیش‌بینی کاربردهای این دستگاه در کارهای آمارگیری و نظامی باعث گردید تا این دستگاه سریع‌تر ساخته شود و مراحل تکامل را به سرعت طی نماید.

روایت‌های گوناگونی در مورد ساخت اولین کامپیوتر وجود دارد و بعضی چرتکه و خط‌کش محاسبه و وسایل مشابه دیگری را نیز از انواع اولیه کامپیوتر می‌دانند اما اولین کامپیوتر رقمی همه منظوره به نام اینیاک (ENIAC - Electronic Numerical Integrator And Calculator) در سال ۱۹۴۶ توسط یک تیم به سرپرستی اکرت و ماچلی از دانشگاه پنسیلوانیا ساخته شد. همچنین کامپیوترهای ادواک (EDVAC - Electronic Discrete Variable Computer) و یونیواک (Universal Automatic Computer) نیز توسط همین گروه طراحی و ساخته شد. قبل از این کامپیوتر در سال ۱۹۳۷ هاوارد ایکن از دانشگاه هاروارد، مارک وان را طراحی نمود که در سال ۱۹۴۴ توسط موسسه IBM (International Business Machine) ساخته شد و در اختیار دانشگاه هاروارد قرار گرفت. این دستگاه ماشین حساب مکانیکی عظیمی بود که قادر بود یک سلسله عملیات طویل حساب و منطق را انجام دهد. این ماشین که دارای واحدهای

حافظه، کنترل و محاسبه بود، دستورالعمل‌ها را به صورت کد از نوار منگنه کاغذی دریافت می‌کرد و نتیجه را بعد از انجام عملیات طبق دستورهای داده شده به آن در اختیار اپراتور قرار می‌داد. البته سرعت مارک وان به علت وابستگی به اجزا الکترومکانیکی بسیار کند بود.

در ساخت اینیاک حدوداً از ۱۸۰۰۰ لامپ خلاء استفاده شده بود و برای انرژی مصرفی لامپ‌ها و همچنین دستگاه‌های تهویه و خنک‌کننده، حدود ۱۵۰ KW انرژی الکتریکی مصرف می‌شد و دارای سطحی ۱۶۰ متر مربع اشغال می‌کرد و دارای ۳۰ تن وزن بود. این دستگاه قادر بود ۵۰۰۰ جمع و ۲۵ ضرب را در مدت یک ثانیه انجام دهد. به این دسته از کامپیوترها، کامپیوترهای نسل اول گفته می‌شود. از مشخصه‌های بارز این نسل می‌توان به موارد زیر اشاره کرد:

۱. سرعت عمل آنها حدود یک هزارم ثانیه بود.
۲. حافظه آنها دارای ظرفیت ۲۰۰ تا ۴۰۰۰ کلمه بود.
۳. دارای کاربردهای ویژه تک مدل بودند.
۴. کلیه برنامه‌ها به زبان ماشین نوشته می‌شد.
۵. در آنها از لامپ خلاء به عنوان حافظه استفاده می‌شد.
۶. دارای وزن و حجم زیادی بودند.
۷. مصرف برق در آنها بسیار زیاد بود.

در اوایل دهه ۱۹۵۰ با اختراع ترانزیستور و استفاده از آن به عنوان حافظه، تغییرات عمده‌ای در کامپیوترها ایجاد کرد که به عنوان نسل دوم کامپیوترها شناخته می‌شوند. اختراع ترانزیستور باعث کوچک شدن کامپیوتر و سبکتر شدن آنها گردید و مصرف برق در آنها بشدت کاهش پیدا کرد. کامپیوترهای نسل دوم در واقع اولین کامپیوترهایی بودند که غیر از دانشگاه‌ها و موسسات تحقیقاتی در موسسات دولتی و شرکت‌های خصوصی برای امور غیر علمی نیز به کار گرفته شدند. اولین کامپیوتری که در ایران نصب گردید از نسل دوم و مدل IBM 1620 بود که در سال ۱۳۴۱ در کنسرسیوم نفت تهران به کار گرفته شد و همچنین سرشماری سال ۱۳۴۵ نیز با استفاده از کامپیوترهای همین نسل، مدل IBM 1401 انجام گردید.

از مشخصه‌های بارز این نسل می‌توان به موارد زیر اشاره کرد:

۱. از ترانزیستور در آنها استفاده شد.
 ۲. سرعت عمل آنها حدود یک میلیونیم ثانیه بود.
 ۳. ظرفیت آنها حدود ۳۰,۰۰۰ کلمه بود و حافظه‌های کمکی نیز در این نسل بوجود آمدند.
 ۴. دارای کاربردهای عمومی یا همه منظوره بودند.
 ۵. زبانهای برنامه‌نویسی آنها، فوق‌العاده ساده‌تر شد.
 ۶. دارای حجم بسیار کمتری بودند.
 ۷. از حلقه‌های کوچک مغناطیسی به عنوان حافظه در آنها استفاده می‌شد.
- ساخت IC (Integrated Circuits) یا مدارات مجتمع باعث تولید کامپیوترهای نسل سوم گردید. در اوایل دهه ۱۹۶۰ اولین کامپیوتر از این نسل با نام IBM 360 با هزینه معادل ۵ میلیارد دلار ساخته شد که بزرگترین پروژه ملی بخش خصوصی تا آن تاریخ به شمار می‌رفت. جدیدترین تحول در تکامل این نسل، ساخت وسایل ضبط اطلاعات با قابلیت دسترسی مستقیم بود و کاربران می‌توانستند در کسری از ثانیه به هر یک از اجزاء اطلاعات ذخیره شده در یک مجموعه عظیم اطلاعاتی دسترسی داشته باشند. مهمترین مشخصات کامپیوترهای نسل سوم عبارتند از:

۱. پیشرفتهای سخت‌افزاری

(الف) تقلیل حجم دستگاه‌ها و اجزای آنها

(ب) افزایش ظرفیت حافظه به چندین برابر قبل

(ج) استفاده از دستگاه‌های واسطه ذخیره‌سازی با قابلیت دسترسی مستقیم

(د) قدرت ارتباط با نقاط دور و متعدد

۲. پیشرفتهای نرم‌افزاری

(الف) هماهنگی بیشتر با سخت‌افزار

(ب) هماهنگی بیشتر با سیستم عامل

ج) پیشرفت در زبانهای برنامه نویسی و به کارگیری زبانهای سطح بالا

۳. عملیات و بهره برداری

الف) استفاده از روشهای پردازش مستقیم

ب) اجرای هم زمان چند برنامه با یکدیگر

تقسیم بندی و تفکیک نسل های کامپیوتری تا نسل سوم، به دلیل تغییرات عمده در پیشرفت و تکامل کامپیوتر در هر نسل به سهولت صورت می گرفت. ولی در سایر موارد اختلاف سلیقه زیادی وجود دارد. در اوایل سال ۱۹۷۰ تکنیکهای جدیدتری در ساخت و بهره برداری از کامپیوترها به کار برده شد که بسیاری از دست اندرکاران آن را نسل چهارم نامیدند و کامپیوترهایی که در حال حاضر از آنها استفاده می کنیم، از همین دسته می باشند. مهمترین تغییرات در ساخت افزار کامپیوترهای نسل چهارم به کار گرفتن مدارهای مجتمع با تراکم بسیار بالا است. همچنین در این نسل از حافظه ها، مدارهای میکروپروسور، سیستم محاوره ای، پردازش با ارتباط مستقیم و شبکه های کامپیوتری استفاده گردید.

کامپیوترهای نسل پنجم که ایده آن اولین بار توسط ژاپنی ها در سال ۱۹۸۰ مطرح شد، پیشنهاد ساختن کامپیوترهایی را دارد که: بتوانند بیاموزند، استدلال کنند و تصمیم بگیرند. به عبارت دیگر رفتاری داشته باشند که معمولاً در حوزه منطق و استدلال خاص از آنها قرار دارند و هوشمندانه کار کند. مسابقات ربانیک که در گوشه و کنار دنیا شاهد برگزاری آن هستیم در این راستا و ایجاد انگیزه در پژوهشگران برای کار و فعالیت بیشتر در این زمینه برگزار می گردد.