

# داده‌پردازی برای کتابداران و اطلاع‌رسانان

• حسین پاشایی‌زاده

عضو هیأت علمی دانشگاه پیام نور

این کتاب در پنج فصل تدوین شده است. فصل اول به عناصر پایگاه اطلاعاتی می‌پردازد و در آن اصول ضبط‌نویسه‌ها در رایانه و اجزای متشکله یک پایگاه اطلاعاتی شرح داده می‌شود. برای اینکه رایانه بتواند داده‌های مورد نظر را در خود ذخیره و مورد پردازش قرار دهد، لازم است داده‌ها بصورت رقمی‌های دودویی در آورده شوند و معمولاً برای این کار از اعداد صفر و یک (۰ و ۱) استفاده می‌شود. در این فصل انواع کدگذاری‌ها، مانند کدگذاری با استفاده از هفت بیت یا کد اسکی (ASCII) که در آن از ۱۲۸ کد برای نمایش‌نویسه‌ها استفاده می‌شود؛ کد هشت‌بیتی شرکت آی بی ام با نام اِی‌بی‌سی‌دی‌سی (EBCDIC) که در این روش کدگذاری نمایش اعداد به یکی از دو حالت فشرده و غیر فشرده صورت می‌گیرد؛ کد استاندارد بین‌المللی که در آن ۱۲۸ کد اول جدول این کدگذاری را به ارقام صفرتا نه و حروف بزرگ و کوچک الفبای زبان انگلیسی و نویسه‌های کنترلی و نشانه‌های خاص اختصاص داده است و ۱۲۸ کد بقیه نیز برای زبان‌دیگر کشورها در نظر گرفته شده است تا هر کشور مطابق با نیاز خود به کدگذاری بپردازد؛ کدگذاری با کد اسکی هشت‌بیتی که یک کد استاندارد آمریکایی است و به جز دو مورد، مشابه کد استاندارد بین‌المللی است، تشریح شده است.

نویسنده در ادامه فصل به کد واحد استاندارد، منطقه و انواع مناطق (حوزه)، فرق بین داده و اطلاع، پیشینه و انواع آن، بایگانی و انواع آن، مهمترین عملیات در پایگاه‌های اطلاعاتی و نهایتاً عناصر متشکله بایگانی واژه‌ها یا نمایه مقلوب پرداخته است. نویسنده در نگارش این فصل از قسمتهایی از فصل «کلیات و مفاهیم» کتاب «مفاهیم و روشهای ذخیره و بازیابی اطلاعات در نظامهای کامپیوتری کتابخانه‌های ایران» (دیانی و داوریانه، ۱۳۷۸) و نیز قسمتهایی از بخش اول کتاب «ذخیره و بازیابی اطلاعات» (صادقی، ۱۳۷۹) با

• داده‌پردازی برای کتابداران و اطلاع‌رسانان، اقتباس و ترجمه: دکتر محمدحسین دیانی. مشهد: انتشارات کتابخانه رایانه‌ای، ۱۳۸۱. چاپ اول، وزیری، ۱۲۹ ص، شمیز.

کتاب حاضر به ساختار و عناصر پایگاه اطلاعاتی و بایگانی (فایل) می‌پردازد و در آن مباحثی چون کدگذاری، انواع بایگانی، انواع جستجو و عملگرها مورد بحث و بررسی قرار می‌گیرند. کتاب با انگیزه تامین منبع درسی مقطع کارشناسی ارشد برای درس «داده‌پردازی» تهیه گردیده است و در تهیه آن از فصول مختلف کتابهای دیگر اقتباس شده و یا فصلهایی از آن کتاب‌ها ترجمه شده و به این کتاب افزوده شده است. نویسنده کتاب، دکتر محمدحسین دیانی، استاد رشته کتابداری و اطلاع‌رسانی دانشگاه فردوسی مشهد هستند و تا به حال آثار زیادی، اعم از کتاب تالیفی و ترجمه‌ای، مقاله و طرح تحقیقی، از ایشان چاپ و منتشر گردیده است.

همانگونه که مولف محترم نیز در مقدمه کتاب اشاره کرده است با توسعه فناوریهای اطلاعاتی و ارتباط تنگاتنگ بین علم اطلاع‌رسانی و علوم رایانه، کتابداران و اطلاع‌رسانان مجبورند دامنه دانش خود را در این علوم گسترش دهند تا با ملزومات دانش رایانه که بهینه‌سازی دانش اطلاع‌رسانی را امکانپذیر می‌سازد، آشنا شوند. مولف معتقد است که نمی‌توان مستقیماً براساس متونی که در حوزه دانش رایانه‌ای و برای رایانه کاران به نگارش در آمده است به آن قسمت از دانش مورد نیاز خود دست یافت. فلذا می‌بایست راهی میانبر انتخاب شود و هدف، تمرکز بر بخشهایی از دانش رایانه‌ای است که در اطلاع‌رسانی کاربرد اساسی دارد و نیز گلچین دانشهای مقدماتی آنها برای ارائه در قالب متونی است که زبان آن را، هم مدرس کتابداری و هم دانشجوی کتابداری درک کند.

می‌رود. نویسنده معتقد است که همه نظام‌های مدیریت بانک‌های اطلاعاتی که برای رایانه‌های شخصی طراحی شده‌اند، بر اساس فناوری یارانه‌ای شکل گرفته‌اند. وی سه دلیل برای ارجحیت این فناوری بر فناوریهای سلسله‌مراتبی و شبکه‌ای ذکر می‌کند:

۱. مدل رابطه‌ای به لحاظ مفهومی به ساده‌ترین شکل اداره می‌شود.

۲. دارای بیشترین انعطاف است.

۳. در کمترین زمان، نظام‌های کاربردی قابل استفاده‌ای را در اختیار کاربران، حتی کاربران تازه کار، قرار می‌دهد. در این فصل جدولی ارائه شده است که به درک بیشتر مطالب کمک می‌کند.

در فصل پنجم عملیات ششگانه‌ای که بر روی بایگانیها (فایله‌ها) انجام می‌شود مورد بحث و بررسی قرار می‌گیرد. در طراحی هر نظام بایگانی دو هدف اصلی از اولویت برخوردارند. این دو هدف عبارتند از (۱) سرعت عملیات در ذخیره و بازیابی و (۲) صرفه‌جویی در مقدار فضایی که در حافظه برای جای‌دهی داده‌ها لازم است. نویسنده بر این باور است که دستیابی به این دو با مراعات ضابطه‌های زیر امکانپذیر است:

- امکان دستیابی سریع برای ذخیره‌سازی و بازیابی داده‌ها وجود داشته باشد.

- بهنگام سازی عملیات ساده باشد تا اطلاعات با کمترین هزینه روز آمد شوند.

- نگهداری نظام از سهولت لازم برخوردار باشد.

- نظام از قابلیت اطمینان بالایی برخوردار باشد.

عملیات ششگانه‌ای که بر روی بایگانیها انجام می‌شود عبارتند از:

۱. واکنشی پیشینه

۲. بازیابی پیشینه بعدی

۳. درج پیشینه تازه (یا روزآمدسازی با درج)

۴. تغییر در پیشینه‌های موجود (روزآمدسازی)

۵. خواندن تمام بایگانی

عنوان «شیوه‌های نمایش داده‌ها و اطلاعات» اقتباس کرده است.

فصل دوم ساختار بایگانی (فایل) نام دارد و ترجمه کامل دو فصل از کتاب Flynn An Introduction to Information Science نوشته Roger R. با عنوان‌های Fglc Structure و Inverted File است. خواننده در این فصل با سه نوع از متداول‌ترین ساختار برای سازماندهی داده‌ها در فایلها (ترتیبی، دستیابی تصادفی و ترتیبی نمایه شده) آشنا می‌شود. همچنین در این فصل، جستجوی بایگانی، روزآمدسازی بایگانی، جستجوی قطعه‌ای و جستجوی دودویی و روشهای بهینه‌سازی آنها، روش‌های برخورد با «سرریز» و مطالب دیگری به طور مختصر بیان می‌شوند.

بایگانی مقلوب عنوان فصل سوم کتاب است. در این فصل به ساختاری از بایگانی پرداخته می‌شود که بر محتوای داده‌های مربوط به کلیدواژه تاکید دارد. بایگانی مقلوب نیز همانند ساختار ترتیبی نمایه شده متشکل از دو یا تعداد بیشتری بایگانی است با این تفاوت که خصلت نمایه‌ای هر کدام متفاوت است. روشهای سنتی ذخیره‌سازی داده‌ها در بایگانی، ایجاد بایگانی داده‌ها، مقلوب کردن بایگانی، جستجوی بایگانی، کارنمای فرآیند، افزایش دادن نقاط دسترسی، قابلیت انعطاف، ذخیره در برابر پردازش، ترکیب پرسشها با استفاده از عملگرهای بولی (NOT, OR, AND)، استفاده از دیاگرام ون برای بازنمون تصویری از کل پایگاه اطلاعاتی یا بخشهایی از آن و سایر امکانات جستجو مثل مجاورت واژه‌ها، جستجوی واژه‌ها هم ریشه، عملگرهای رابطه‌ای ( $\rightarrow$ ) و غیره از موضوعات مطرح شده در این فصل را تشکیل می‌دهند.

فصل چهارم به ساختار بایگانی در پایگاه اطلاعاتی رابطه‌ای می‌پردازد. این نوع ساختار بایگانی که بویژه در اوپک‌ها (OPACs) کاربرد دارد تحت نرم‌افزار نظام پایگاه مدیریت رابطه‌ای ساخته می‌شود. اینگونه نظام به ذخیره‌سازی اطلاعات در مجموعه‌ای از جداول بهم پیوند داده شده نیاز دارد. اینگونه فایل برای جلوگیری از تکرارهای ناخواسته و کاستن از عدم یکدستی در چند بایگانی بکار

## ۶. بازسازماندهی

در ادامه این فصل پس از شرح مختصری درباره هر کدام از موارد بالا به تعریف پایگاه اطلاعاتی و انواع آن پرداخته شده است. در تعریف پایگاه اطلاعاتی آمده است: «پایگاه اطلاعاتی، مجموعه‌ای است از داده‌های ذخیره شده (درباره موجودیتهای یک محیط عملیاتی و ارتباطات آنها) به صورت مجتمع و مبتنی بر یک ساختار و تعریف شده به طور صوری با استفاده از یک زبان خاص برنامه‌نویسی، با حداقل افزونگی، تحت کنترل متمرکز، مورد استفاده یک یا چند کاربر، بطور اشتراکی و همزمان». نویسنده انواع پایگاه‌های اطلاعاتی را به لحاظ تفاوت داده‌هایی که در خود جای می‌دهند به دو دسته: الف. پایگاه‌های اطلاعاتی مرجع - کتابشناختی یا ارجاعی وب. پایگاه اطلاعاتی منبع تقسیم می‌کند. پایگاه‌های اطلاعاتی کتابشناختی، اطلاعات اصلی مربوط به یک اثر - کتاب یا مقاله یا... - را ارائه می‌کند و عموماً کلید واژه‌های موضوعی، شماره طبقه‌بندی و در بسیاری موارد چکیده یا خلاصه مدرکی را نیز در بردارند ولی پایگاه اطلاعاتی منبع عمدتاً در زمینه‌های اقتصادی، بازرگانی یا تجاری و مالی ساخته می‌شوند. پیشینه‌های این گونه پایگاه‌های اطلاعاتی استفاده کنندگان را به کتابها یا موارد مشابه ارجاع نمی‌دهند، بلکه اطلاعات کامل را به فوریت در اختیار آنان قرار می‌دهند.

## نقد و نظر

رشته کتابداری و اطلاع‌رسانی در ایران سابقه چندانی ندارد و از اینرو از کمبود منابع درسی پرمحتوا به زبان فارسی رنج می‌برد. بنابراین هر گونه تلاشی در جهت پر بار کردن فنون درسی و کمک‌درسی این رشته از طرف هر فردی که باشد ستودنی است و دکتر محمدحسین دبانی بالاخص در این زمینه تلاشهای زیادی بعمل آورده است که جا دارد تقدیر مضاعفی از ایشان بعمل آید. با این حال این کتاب دارای نواقص و اشکالات اساسی است که نمی‌توان بسادگی از کنار آن عبور کرد بویژه اینکه کتاب توسط فرد

صاحب‌نامی همچون نویسنده محترم، تهیه و منتشر شده باشد.

اول اینکه کتاب پر از غلطهای چاپی است. این ایراد زمانی برجسته می‌نماید که دانسته شود مسئول انتشارات کتابخانه رایانه‌ای (ناشر کتاب) خود مولف است ولی گویا کنترل لازم را بر روی مراحل مختلف چاپ و بویژه نمونه‌خوانی انجام نداده‌اند. در زیر به اشکالات چاپی کتاب اشاره می‌شود:

ص ۲ - آفریکا/ ص ۲ - جدول (۱.۱) کذاسلکی هفت بیتی / ص ۳ - چهار بیت از این بیت‌های بیت‌های عددی (تکرار کلمه) / ص ۳ - از فام/ ص ۶ - متوزان / تعبیر / ص ۸ - مدم / بابت [بایت] / ص ۹ - جدول (۴.۱) / ص ۱۰ - ۱ ر / ص ۱۳ - بحالی / ص ۱۶ - بلاک (Block) / ص ۱۷ - پیشینه‌ها/ نشاندهن / برنامه / ص ۱۹ - داری/ ص ۲۳ - پایگاه‌های اطلاعاتی / ص ۲۵ - نظر / نرم‌افزار / ارزیابی مجموعه / ص ۲۹ - بایگانی نامنظم / ص ۳۱ - در یک بایگانیهای ترتیبی / متذکر / ص ۳۲ - سوابق / ذخیره / فراهم آورد/ ص ۳۳ - کلید / بازنظمی‌دهی / جستجو / ص ۳۴ - عددی قابل توجهی برای صرفه‌جویی / ص ۳۴ - در بردارند/ جمع این دو عدد / ص ۳۹ - بنا بر اساس / برای جستجو دو حالتی ص ۴۵ - استغاده / ص ۴۶ - مقال / تعداد اندکی سابق / جستجوهای گزینشی / جبران / ص ۴۷ - بازیابی گرینشی / ص ۴۸ - ذخیره‌سازی / ص ۵۰ - تلاق / ص ۵۱ - ذخیره / ص ۵۲ - خصوصیا/ ص ۵۴ - مطالعه الگوریتم‌های گدگذاری / الگوریتم گدگذاری / الگوریتمی دیگری / ص ۵۷ - بایگانیهای دسترسی / ص ۵۸ - میسر است است/ ص ۶۵ - نمایه‌فهرستبرگه‌های در برگه‌دان / ص ۶۶ - رویکرد / نصادفی/ ص ۶۷ - ساختارها بایگانی / به دسترس می‌آیند/ ص ۶۸ - فراری / ص ۶۹ - رزرو و بلیط خط هوایی / نیر/ بیشترک / ص ۷۰ - بایگانی / ص ۷۳ - بازیاند / ص ۷۸ - متوسط قیمت / ص ۹۱ - به مثالی دیگری / تجربه کاری / ص ۹۴ - ترکیب نتایج / ص ۹۶ - همبرگر و تخم‌مرغ / ص ۱۰۰ - جبرل بولی / قرمز و ارزان یا متوسط قیمت/ ص ۱۰۱ - تقدم و تاخیر / ص ۱۰۲ - مجموعه جهانی / افزون بر

ص ۷۰ - در یکبار کاربرد، شغل‌های اورژانس، یا تقاضاهای یگانه، کارایی ساختار بایگانی بیشتر دقیق بقدر کافی توجیه کننده اختصاص زمان اضافی برای تهیه وجود ندارد.

ص ۹۳ - دارای تجربه در یکی بودن مانع تجربه داشتن در کار دیگر نیست. لذا، می‌توانیم ارزشهای داده‌ای همپوشان در منطقه تجربه‌کاری باشیم.

ص ۱۰۶ - ... نتیجه بازیابی اطلاعات غالباً به یکی از این شکلها، هر ترکیبی، از این سه باز نمون متفاوت اعلام می‌شود.

ص ۱۲۸ - با توجه به این ضوابط، مشخص است که در هر پایگاه اطلاعاتی شش نوع کار که در زیر فهرست شده‌اند، روی بایگانیها انجام می‌دهد.

ص ۱۳۳ - ... در بحث داده بازیابی مبتنی بر پردازش داده‌ها مورد توجه می‌باشند.

کتاب فاقد ساختار مشخصی است که به این علت می‌تواند باشد که مطالب آن از فصول مختلف کتابهای مجزا تهیه شده است. همچنین مثالهایی که در آن استفاده شده اغلب از حوزه‌هایی غیر از حوزه کتابداری و اطلاع‌رسانی است. نکته آخری که توجه خواننده رابه خود معطوف می‌کند منابع مورد استفاده برای تهیه این کتاب است.

اول اینکه صفحات مورد استفاده از منابع اول و دوم آخر کتاب ذکر نشده است. در منبع چهارم علاوه بر اینکه صفحات مورد استفاده ذکر نگردیده نحوه ارائه اطلاعات نیز غلط است چرا که اسم نویسنده دوم می‌بایست بعد از نویسنده اول می‌آمد و نیز تکرار نام نویسنده اول بعد از عنوان کتاب صحیح نیست. در منابع پنجم، ششم، هفتم و هشتم فارسی و نیز منبع دوم لاتین صفحات مورد استفاده قید نشده است. همچنین با توجه به تغییرات سریع صورت گرفته در حوزه ذخیره و بازیابی منابع لاتین مورد استفاده نسبتاً کهنه و قدیمی به نظر می‌رسند (1987, 1985). امید که شاهد انتشار کتابهای پربارتری از مولف محترم باشیم.

تصویرگردن جامعه ص ۱۰۵ - نظریه مجموعه / ص ۱۰۶ - نتایج / کاربر / ص ۱۰۸ - بازیابی / ص ۱۱۰ - هستند / ص ۱۱۲ - نرم افزار / نظام‌هایی / ص ۱۱۳ - زبانی یا، تاریخی یا مکانی / New technology / ص ۱۱۵ - نظام / کاربردهای / ص ۱۱۶ - تعبیر / ص ۱۱۷ - ترکیبها ممکن / ص ۱۱۸ - در جدول (۴ - ۲) / ص ۱۲۰ - کتابچه / ص ۱۳۰ - پاسخ / ص ۱۳۲ - موارد یا مواردی / ص ۱۳۳ - عملیاتی / ص ۱۳۸ - هستند /

همچنین در این کتاب برخی جملات گنگ و نامفهوم مشاهده می‌شود.

ص ۳۵ - ... مشاهده می‌کنیم که صرفه جویی قابل ملاحظه‌ای که تقریباً برابر ۸۰۰ جستجو در جستجوی بدترین حالت میسر شده است. ص ۵۷ - ... با حداقل زمان جستجو یک دسترسی، و مورد متوسط و بدترین نه بیشتر از یک، دسترسی تصادفی فراهم آورنده انگیزه‌ای اغواکننده است.

ص ۴۶ - ... مثلاً، کار عادی خاص برنامه جستجوی دو حالتی ممکن است سه برابر کار عادی خاص برنامه جستجوی متوالی زمان می‌برد.

ص ۴۶ - مغایر با بازیابی گزینش سوابق، بازیابی «کامل» است. ص ۴۸ - از این داده‌ها برای تعیین محل ذخیره‌سازی به شرح زیر استفاده می‌کنیم. به جایگاه هر حرف در مجموعه حروف الفبای انگلیسی بترتیب شماره‌ای اختصاص می‌دهیم و مجموعه اعدادی اختصاص داده شده به حروف یک نام، محل ذخیره را تعیین می‌کنیم. ص ۵۰ - بنابراین، تعداد کمتری سابقه قابل ذخیره شدن هستند. اما از هزینه‌های مکانهایی که به حوزه سرریز تخصیص داده می‌شود، اما به دلیل نبود سابقه‌ای با مشخصات آنها بلا استفاده بمانند صرفه‌جویی می‌کنیم.

ص ۶۳ - نظر این است که نوعی نمایه‌سازی داشته باشیم که نمایه‌ای برای نمایه مربوط به داده‌ها باشیم.

ص ۶۹ - ۷۰ - یک قاعده سردستی خوب این است که ...