

# ذخیره و بازیابی اطلاعات

دکتر محمد حسن زاده

تهران

۱۳۹۷



سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها (سمت)  
پژوهشکده تحقیق و توسعه علوم انسانی

## فهرست مطالب

صفحه	عنوان
سیزده	فهرست شکل‌ها
شانزده	فهرست جدول‌ها
۱	پیشگفتار
بخش اول: ذخیره اطلاعات	
۹	فصل اول: اطلاعات و تصمیم
۹	مقدمه
۹	تعریف اطلاعات
۱۰	زیست‌بوم اطلاعات
۱۱	اطلاعات و نظریه سیستمی
۱۳	اطلاعات و عدم قطعیت
۱۴	اطلاعات و آنتروپی
۱۶	اطلاعات و ارتباط آن با سطوح گوناگون مدیریتی
۱۷	نظام‌های اطلاعات مدیریتی
۱۷	اطلاعات و تصمیم‌های فردی
۱۸	خودآزمایی
۲۰	فصل دوم: مفاهیم بنیانی
۲۰	مقدمه
۲۰	مسیر رشد بازیابی اطلاعات
۲۱	دوره قدیم
۲۳	دوره جدید

صفحه	عنوان
۲۴	وب، انقلابی بزرگ
۲۵	ماهیت بازیابی اطلاعات
۲۶	تعاریف بازیابی اطلاعات
۲۹	کاربردها و کارکردهای بازیابی اطلاعات
۳۰	نکات مهم برای متخصصان علم اطلاعات
۳۱	بازیابی اطلاعات و ارزش افزوده
۳۲	خودآزمایی
۳۳	<b>فصل سوم: فناوری پایگاه‌های اطلاعاتی</b>
۳۳	مقدمه
۳۳	تعریف پایگاه اطلاعاتی
۳۴	انواع پایگاه‌ها
۳۶	اجزاء و عناصر یک پایگاه
۳۹	مدل داده
۴۱	انواع مدل داده
۴۵	نظام مدیریت پایگاه داده
۴۵	وظایف نظام مدیریت پایگاه داده
۴۶	تراکنش
۴۷	اجزاء نظام مدیریت پایگاه داده
۴۹	انواع نظام‌های مدیریت پایگاه داده
۵۳	معماری پایگاه داده
۵۴	سطح خارجی
۵۵	سطح مفهومی
۵۵	سطح داخلی
۵۷	استقلال داده
۵۷	زبان برنامه‌نویسی پایگاه داده
۵۸	زبان جاوا
۵۹	زبان اس کیوال
۶۲	خودآزمایی

صفحه	عنوان
۶۳	فصل چهارم: قالب‌ها و استانداردها
۶۳	مقدمه
۶۳	کتابداران و استانداردهای کتاب‌شناختی
۶۵	استانداردها و دنیای کتاب‌شناختی
۶۶	سازمان‌ها و نهادهای مرتبط با استانداردهای کتاب‌شناختی
۶۸	استانداردهای کتاب‌شناختی
۶۹	فرمت مارک
۷۳	مارک ایکس ام ال
۷۴	هسته دویلین (دویلین کور)
۷۵	فرمت‌های توصیف منابع
۷۶	فرمت گسترش‌یافته تاریخ/ زمان (EDTF) تعریف جامع تاریخ / ...
۷۷	استاندارد توصیف مستند ابر داده (MADS)
۸۱	اقدام چهارچوب کتاب‌شناختی
۸۴	استانداردهای خاص کتابخانه‌های دیجیتال
۹۰	پروتکل‌های بازیابی منابع اطلاعاتی
۹۰	سی کیو ال (زبان پرسش زمینه‌ای)
۹۱	اس آر یو / اس آر دبلیو
۹۶	ملزومات کارکردی پیشینه‌های کتاب‌شناختی
۱۰۰	خودآزمایی
۱۰۲	فصل پنجم: بازنمایی اطلاعات
۱۰۲	مقدمه
۱۰۳	بازنمایی چیست؟
۱۰۵	انواع بازنمایی اطلاعات
۱۰۷	جایگاه بازنمایی در بازیابی اطلاعات
۱۰۸	چالش‌های بازنمایی در محیط اینترنت
۱۱۱	پیش‌همارایی و پس‌همارایی
۱۱۲	نمایه چیست؟
۱۱۴	دیدگاه‌های نظری در خصوص نمایه‌سازی
۱۱۵	ویژگی‌های نمایه‌ساز

صفحه	عنوان
۱۱۶	قواعد نمایه‌سازی
۱۱۷	عمق نمایه‌سازی
۱۱۸	فرایند نمایه‌سازی
۱۲۰	خودآزمایی
۱۲۱	<b>فصل ششم: کنترل واژگانی و مستندسازی</b>
۱۲۱	مقدمه
۱۲۲	اهداف و کارکردهای کنترل واژگان
۱۲۳	سرعنوان‌های موضوعی
۱۲۳	اصطلاح‌نامه
۱۲۵	ساختار اصطلاح‌نامه
۱۲۶	مراحل ساخت اصطلاح‌نامه
۱۲۷	نظام اسکاس
۱۲۸	ساختار اسکاس
۱۳۱	نمونه‌ای از کاربرد اسکاس در اصطلاح‌نامه‌ها
۱۳۳	نرم‌افزار اصطلاح‌نامه
۱۳۶	خودآزمایی
<b>بخش دوم: بازیابی اطلاعات</b>	
۱۳۹	<b>فصل هفتم: مدل‌ها و الگوریتم‌های بازیابی</b>
۱۳۹	مقدمه
۱۳۹	تعریف بازیابی
۱۴۰	بازیابی جانشین مدرک در مقابل بازیابی خود مدرک (از واژگان به تمام متن)
۱۴۱	فرایند بازیابی
۱۴۲	مدل‌های بازیابی اطلاعات
۱۴۵	مدل‌های بولی
۱۴۸	مدل‌های برداری
۱۵۵	مدل‌های احتمالاتی
۱۵۹	مدل‌های ساختاریافته
۱۶۲	مدل‌های مروری

صفحه	عنوان
۱۶۵	جمع‌بندی مدل‌های بازیابی اطلاعات
۱۶۸	خودآزمایی
۱۷۰	فصل هشتم: نیاز اطلاعاتی، رفتار اطلاع‌یابی و نظام‌های بازیابی اطلاعات
۱۷۰	مقدمه
۱۷۱	نیاز اطلاعاتی کاربران
۱۷۲	مراحل بیان نیاز اطلاعاتی
۱۷۵	رفتارهای اطلاعاتی
۱۷۸	مدل‌های رفتار اطلاعاتی
۱۸۵	رویکرد شناختی در بازیابی اطلاعات
۱۸۶	فرایند شناخت
۱۸۷	ماهیت ربط شناختی
۱۸۸	خودآزمایی
۱۹۰	فصل نهم: قضاوت ربط در بازیابی اطلاعات
۱۹۰	مقدمه
۱۹۰	تعریف ربط
۱۹۱	عوامل مؤثر بر ربط
۱۹۳	محاسبه ربط در بازیابی
۱۹۳	ارزیابی ربط
۱۹۴	ربط شناختی
۱۹۷	ادراک و ربط
۱۹۸	توجه و ربط
۲۰۰	قضاوت و ربط
۲۰۱	درک شهودی ربط و فرایند شناخت
۲۰۲	ضرورت توجه متخصصان اطلاع‌رسانی به ارتباط شناخت و ربط
۲۰۴	خودآزمایی
۲۰۵	فصل دهم: رابط کاربر
۲۰۵	مقدمه
۲۰۶	جایگاه رابط کاربر

صفحه	عنوان
۲۰۸	ویژگی‌های رابط کاربر
۲۰۹	مراحل رشد رابط کاربر
۲۰۹	انواع رابط کاربر
۲۱۴	ملاحظات طراحی رابط کاربر
۲۱۶	طراحی رابط کاربر
۲۱۹	ارزیابی رابط کاربر
۲۱۹	فرایند ارزیابی
۲۲۱	روش‌های ارزیابی
۲۲۷	خودآزمایی
۲۲۹	<b>فصل یازدهم: بازیابی اطلاعات در اینترنت و وب</b>
۲۲۹	مقدمه
۲۲۹	تاریخچه مختصر اینترنت
۲۳۳	ویژگی‌های اینترنت
۲۳۴	ساختار اینترنت
۲۳۵	راهنماها
۲۳۷	موتور جستجو
۲۴۲	ابرموتورهای جستجو
۲۴۶	خودآزمایی
۲۴۸	<b>فصل دوازدهم: ارزیابی و سنجش نظام‌های ذخیره و بازیابی</b>
۲۴۸	مقدمه
۲۴۸	ضرورت‌های ارزیابی
۲۵۱	۱. تعیین میزان آمادگی
۲۵۸	۲. سنجش کارایی
۲۵۹	۳. سنجش اثربخشی
۲۶۰	۴. سنجش بهره‌وری
۲۶۰	۵. سنجش عملکرد
۲۶۶	جمع‌بندی
۲۶۶	خودآزمایی

صفحه	عنوان
۲۶۸	فصل سیزدهم: بازیابی اطلاعات دانش‌مدار
۲۶۸	مقدمه
۲۶۹	ویژگی‌های بازیابی اطلاعات دانش‌مدار
۲۷۰	نگاشت صرفی و نحوی
۲۷۱	تعامل با کاربر
۲۷۲	ابزارهای واسط
۲۷۸	چهارچوب مفهومی
۲۸۰	سخن پایانی
۲۸۱	خودآزمایی
۲۸۲	فصل چهاردهم: فنون بازیابی اطلاعات
۲۸۲	مقدمه
۲۸۲	نحوه عبارت‌بندی جستجو
۲۸۳	عملگرهای جستجو
۲۸۳	عملگرهای منطقی
۲۸۴	عملگرهای ترکیبی
۲۸۵	انواع جستجو (ساده و پیشرفته)
۲۸۶	اشتباهات معمول در انجام دادن جستجو
۲۸۸	مثال‌هایی از قابلیت‌های جستجو
۲۹۱	خودآزمایی
۲۹۲	فصل پانزدهم: بازیابی اطلاعات چندرسانه‌ای
۲۹۲	مقدمه
۲۹۲	ماهیت و اهداف نظام‌های نمایه‌سازی چندرسانه‌ای
۲۹۳	فیلم
۲۹۴	استفاده‌کنندگان فیلم
۲۹۴	اجزاء اطلاعاتی فیلم
۲۹۶	تصاویر
۲۹۶	اجزاء اطلاعاتی تصاویر
۲۹۷	تصاویر الکترونیکی



صفحه	عنوان
۳۰۰	برنامه‌های رایانه‌ای
۳۰۱	ابزارهای جستجوی تصاویر
۳۰۲	خودآزمایی
۳۰۳	سخن پایانی: حوزه‌های نوین پژوهش در ذخیره و بازیابی اطلاعات
۳۰۳	روند پژوهش‌ها
۳۰۷	آینده ذخیره و بازیابی اطلاعات
۳۰۹	فهرست منابع
۳۱۴	پیوست الف: فهرستی از نرم‌افزارهای ذخیره و بازیابی
۳۱۵	پیوست ب: نمونه‌هایی از موتورهای جستجوی تصاویر بر مبنای محتوا
۳۱۶	واژه‌نامه فارسی - انگلیسی
۳۱۹	واژه‌نامه انگلیسی - فارسی
۳۲۳	نمایه

## پیشگفتار

خداوند سبحان را شاکرم که توفیق حضور در جبهه علم و دانش را به اینجانب ارزانی داشت و یاریگر شد تا بتوانم این اثر را به جامعه علمی و حرفه‌ای تقدیم نمایم. ذخیره و بازیابی اطلاعات به عنوان موضوعی مهم و یکی از درس‌های اساسی در رشته‌های گوناگون از جمله علوم رایانه، علم اطلاعات و دانش‌شناسی بستر مناسبی برای ایجاد نظام‌های کاربرپسند و ارائه خدمات اطلاع‌رسانی و دستیابی به اطلاعات برای عموم مردم فراهم می‌آورد. همه کسانی که در این حوزه به کار پژوهش و آموزش مشغول هستند با ماهیت دوگانه‌ای سروکار دارند. از یک سو، ذخیره و بازیابی اطلاعات فعالیتی همگانی است و عموم مردم از رشته‌ها و قشرهای گوناگون با آن سروکار دارند. از سوی دیگر، حوزه‌ای تخصصی و فنی با پیچیدگی بالاست. حل این معما نیازمند ظرافت خاصی است.

این کتاب برای متخصصان رشته‌های گوناگون که به نوعی با ذخیره و بازیابی اطلاعات سروکار دارند از جمله دانشجویان و دانش‌آموختگان مهندسی کامپیوتر به ویژه مهندسی نرم‌افزار، فناوری اطلاعات، مهندسی سیستم‌ها و نظایر آن مفید خواهد بود؛ اما مخاطب اصلی این کتاب دانشجویان کارشناسی ارشد علم اطلاعات و دانش‌شناسی در گرایش‌های گوناگون‌اند. درس «ذخیره و بازیابی اطلاعات» در قالب دو واحد در گرایش‌های مدیریت اطلاعات و مدیریت کتابخانه‌های دانشگاهی، در مقطع کارشناسی ارشد تدریس می‌شود. علاوه بر آن، درس‌های مرتبط با ذخیره و بازیابی اطلاعات در مقطع دکتری به ویژه در گرایش بازیابی اطلاعات و دانش نیز به صورت گسترده با این مفهوم مرتبط هستند.

این کتاب شامل دو بخش و ۱۵ فصل است. بخش اول، شامل شش فصل، به

مباحث حوزه ذخیره اطلاعات و بخش دوم حاوی نه فصل بعدی، به مفاهیم و فرایندهای بازیابی اطلاعات اختصاص داده شده است. در فصل اول به کلیات مرتبط با نظام‌های ذخیره و بازیابی اطلاعات توجه شده است و مباحثی مانند چستی اطلاعات، زیست‌بوم اطلاعات، نظریه سیستمی، عدم قطعیت، آنتروپی و کارکرد آن در جریان اطلاعات و همچنین نقش اطلاعات در سازمان‌ها و تأثیر آن‌ها بر فرایند تصمیم‌گیری به صورت اجمالی تبیین می‌شود. در تدوین این فصل سعی بر آن بوده است که با متنی روان، نگاه تحلیلی به مباحث معطوف گردد.

در فصل دوم، مفاهیم بنیادین بازیابی اطلاعات مورد توجه بوده است. در این راستا، تاریخچه مختصری از ذخیره و بازیابی اطلاعات ارائه شده و کاربردها و کارکردهای بازیابی اطلاعات تبیین شده است. از آنجایی که وب نقطه عطف در روند شکل‌گیری و تکامل نظام‌های ذخیره و بازیابی اطلاعات به حساب می‌آید، هم در این فصل و هم در فصل‌های دیگر، هر جا که لازم بوده به صورت جدی به این مفهوم پرداخته شده است. بحث‌های پایانی فصل دوم به نکات ضروری برای متخصصان علم اطلاعات و دانش‌شناسی در خصوص ذخیره و بازیابی اطلاعات و ارزش‌آفرینی از طریق بازیابی اطلاعات اختصاص یافته است.

فصل سوم اولین فصل از کتاب است که در آن به مباحث فنی پرداخته شده است. در این فصل، ابتدا پایگاه اطلاعاتی تعریف شده و اجزاء و عناصر یک پایگاه اطلاعاتی از جمله مدل داده و انواع آن به تصویر کشیده شده است و سپس انواع معماری‌های مربوط به پایگاه اطلاعاتی و معرفی برخی زبان‌های برنامه‌نویسی در ادامه این فصل مورد توجه قرار گرفته است.

فصل چهارم کتاب به قالب‌ها و استانداردهای حوزه ذخیره و بازیابی اطلاعات مربوط می‌شود. در این فصل ابتدا جایگاه استاندارد در دنیای ذخیره و بازیابی اطلاعات به ویژه نقطه شروع آن، یعنی حوزه کتاب‌شناختی، مورد بحث قرار گرفته است. در ادامه استانداردهایی مانند هسته دوبلین، انواع مارک‌ها، قالب‌های نوین مانند مودز، متز، تکست ام دی، استاندارد Z39.50 و نظایر آن معرفی شده است. خواندگانی که به پیشرفت‌های نوین در حوزه استانداردهای ذخیره و بازیابی

اطلاعات به ویژه حوزه کتاب‌شناختی علاقه‌مند هستند، این فصل را بسیار مفید خواهند یافت.

فصل پنجم کتاب، درباره بازنمایی اطلاعات است. در این فصل انواع بازنمایی اطلاعات و چستی آن‌ها به همراه چالش‌های بازنمایی اطلاعات و منابع گوناگون اطلاعاتی توضیح داده شده است. مباحث قدیمی مانند همارایی و انواع آن به همراه بحث گسترده‌ای در حوزه نمایه‌سازی در ادامه این فصل آورده شده است. کسانی که می‌خواهند با فرایند نمایه‌سازی و نحوه تعیین عمق آن و قواعد مربوطه آشنا شوند، در این فصل به آن‌ها دست خواهند یافت.

در فصل ششم، آخرین فصل از بخش اول، کنترل واژگانی مورد توجه قرار گرفته است. واقعیت این است که کنترل واژگانی یکی از مشکلات جاری در حوزه ذخیره و بازیابی اطلاعات تلقی می‌شود. با جستجو در موتورهای جستجو با ریزش کاذب بسیار بالا و عدم بازیابی پیشینه‌های مرتبط روبه‌رو هستیم. شاید بخشی از این عملکرد موتورها و عامل انسانی به نمایه‌سازی مناسب محتوایی مربوط باشد، اما بخش مهمی از چالش به نبود کنترل واژگانی و نبود برآورد شفاف از روابط واژگانی حاکم بر جهان منابع اطلاعاتی ارتباط دارد. در این فصل علاوه بر مباحث نظری، فرایند ساخت اصطلاح‌نامه به همراه ابزارها و چهارچوب‌های نوین توضیح داده شده است.

بخش دوم با فصل هفتم کتاب آغاز می‌شود. این فصل به مدل‌ها و الگوریتم‌های بازیابی اطلاعات اختصاص یافته است. به صورت مفصل، توضیحات مربوط به مدل‌های بولی، برداری، احتمالاتی و مروری در این فصل ارائه شده است. خوانندگان این فصل با جامع‌ترین اطلاعات مربوط به مدل‌های بازیابی اطلاعات روبه‌رو خواهند شد. جمع‌بندی مدل‌ها در قالب یک جدول، پایان‌بخش این فصل است.

در فصل هشتم، به صورت ویژه نیاز اطلاعاتی و رفتار اطلاعاتی مورد بحث قرار گرفته است. اولین بحث در این رابطه تعریف نیاز اطلاعاتی و مراحل آن بوده

است. مراحل بیان نیاز اطلاعاتی مخصوص این کتاب از خوداحساسی تا بیان رسمی توضیح داده شده و سپس به شرح انواع رفتارها و مدل‌های تبیین‌کننده رفتار اطلاعاتی پرداخته شده است. تا حد امکان تلاش شده است که از طولانی شدن بحث مربوط به مدل‌ها جلوگیری شود. در پایان این فصل، با مطالعه مفهوم رویکرد شناختی در بازیابی اطلاعات افق جدیدی پیش روی خوانندگان قرار خواهد گرفت.

فصل نهم به قضاوت ربط اختصاص یافته است. تعریف ربط، عوامل مؤثر بر ربط و نحوه محاسبه آن در این فصل توضیح داده شده است. مباحثی مانند ربط شناختی و درک شهودی ربط در ادامه این فصل آورده شده است.

فصل دهم کتاب به رابط کاربر اختصاص یافته است. اهمیت رابط کاربر در نظام‌های ذخیره و بازیابی اطلاعات به بازگویی نیاز ندارد. با وجود این، طراحی رابط کاربر و ارزیابی آن به صورت شگفت‌انگیزی در کتاب‌های ذخیره و بازیابی اطلاعات نادیده گرفته شده است. فرایند ارزیابی رابط کاربر در قالب کاربرپذیری، کاربرپژوهی و دسترس‌پذیری توضیح داده شده است.

فصل یازدهم کتاب به صورت ویژه برای کسانی طراحی شده است که به اینترنت و مسائل آن علاقه‌مند هستند. اینترنت نوع ویژه‌ای از ذخیره و بازیابی را به جامعه علمی و حرفه‌ای تحمیل کرده است. بر همین اساس، در این فصل به صورت جداگانه این موضوع مورد بحث قرار گرفته است. ابتدا تاریخچه بسیار مختصر و مفیدی از اینترنت و عملکرد آن در حوزه ذخیره و بازیابی اطلاعات ارائه شده و در ادامه، ساختار اینترنت و راهنماها و موتورها و ابرموتورهای جستجو توضیح داده شده است. همین‌طور الگوریتم‌های رایج در موتورهای جستجو برای مواجهه با مسائل بهینه‌سازی موتور جستجو در این قسمت ارائه شده است.

فصل دوازدهم به ارزیابی و سنجش نظام‌های ذخیره و بازیابی اطلاعات مربوط است. در این فصل توضیحات جامع و راهکارهای مناسبی برای ارزیابی بازیابی اطلاعات ارائه شده است. مفهوم ارزیابی از منظر آمادگی فنی، آمادگی

سازمانی، سنجش کارایی، عملکرد، بهره‌وری و نظایر آن مورد توجه قرار گرفته است. سنججه‌هایی مانند بازیافت، دقت و اِف نیز به همراه فرمول‌های مربوطه توضیح داده شده است.

در فصل سیزدهم کتاب مفهوم نوینی پیش کشیده شده است. مفهوم بازیابی اطلاعات دانش‌مدار هنوز هم در مراحل اولیه خود قرار دارد و کسانی که در حوزه ذخیره و بازیابی اطلاعات فعالیت می‌کنند، برای دستیابی به نظام‌های دانش‌مدار راه درازی در پیش دارند. در این فصل تلاش شده است ضمن تبیین عناصر، ارکان و مفاهیم مربوط به بازیابی اطلاعات دانش‌مداری، دیدگاهی مناسب در قالب چهارچوب مفهومی برای خوانندگان ارائه شود.

فصل چهاردهم به صورت عملیاتی به فنون بازیابی اطلاعات اختصاص دارد. در این فصل، نحوه عبارت‌بندی مناسب نیاز اطلاعاتی در قالب عبارت جستجو آموزش داده شده است. عملگرهای مورد استفاده برای جستجو جمع‌بندی و ارائه شده است. در ادامه برخی اشتباهات معمول در زمینه جستجوی اطلاعات به خواننده گوشزد شده و مثال‌هایی از قابلیت‌های پرکاربرد جستجو در موتورهای جستجو ارائه شده است.

در فصل پانزدهم کتاب بازیابی اطلاعات چندرسانه‌ای‌ها مورد توجه قرار گرفته است. در این زمینه مواردی مانند بازیابی تصاویر، فیلم و برنامه‌های رایانه‌ای توضیح داده شده است. تنظیمات مربوط به تصاویر الکترونیکی قسمتی از این فصل را به خود اختصاص داده است.

سخن پایانی کتاب به صورت مجمل و کوتاه به افق‌های پیش روی پژوهش‌های حوزه ذخیره و بازیابی اطلاعات اختصاص دارد. در این راستا ابتدا روند کنونی انتشارات در حوزه بازیابی اطلاعات تبیین شده و در ادامه برخی مفاهیمی که در آینده در حوزه ذخیره و بازیابی اطلاعات به آن توجه خواهد شد، مورد تأکید قرار گرفته است.

سعی نویسنده بر این بوده است که تا حد امکان به منابع مورد استفاده، استناد

داده شود و متنی روان پیش روی خوانندگان قرار گیرد. چنانچه خوانندگان محترم مواردی از نقص در هر مورد مشاهده کرده باشند و آن را به اطلاع نگارنده برسانند، با گوش جان پذیرا خواهم بود. کتاب حاضر به پیشنهاد دکتر عبدالرضا نوروزی چاکلی و به قصد رفع خلأ کتابی درسی در مقاطع تحصیلات تکمیلی تدوین شده، لذا از مساعی ایشان و همکارانشان در مدیریت تدوین سازمان سمت برای انتشار آن، تشکر و قدردانی می‌نمایم.

محمد حسن زاده

تابستان ۱۳۹۷

[hasanzadeh@modares.ac.ir](mailto:hasanzadeh@modares.ac.ir)