



سمیه نازی روانی
دانشگاه علوم پزشکی کلان

دکتر نجارحزینی

معاون امور اطلاعات و کتابخانه‌های
معاون امور اطلاعات و کتابخانه‌های

نظام‌های اطلاعات

www.ketab.ir

نادی راوندی، سمیه، ۱۳۶۰-

نظام‌های بازیابی اطلاعات

مولفان سمیه نادی راوندی و نجلا حریری

تهران: نشر کتابدار، ۱۳۹۴

۲۲۰ ص. مصور (کتابدار: ۳۰۹)

کتابنامه

موضوع: ذخیره و بازیابی اطلاعات

موضوع: پایگاه‌های اطلاعاتی

شناسه افزوده: حریری، نجلا، ۱۳۴۰-

رده کنگره: ۱۳۹۴ ن ۶ ز ۶۹۹/۲

رده دیوئی: ۲۵/۰۴

شماره کتابشناسی ملی: ۴۱۵۳۲۱۲



www.ketabdar.org

mohseni@ketabdar.org

عنوان: نظام‌های بازیابی اطلاعات

مولفان: سمیه نادی راوندی و دکتر نجلا حریری

نوبت چاپ: اول، ۱۳۹۴

شمارگان: ۱۰۰۰ نسخه

قیمت: ۱۵۰۰۰ تومان

شابک: ۹۷۸-۶۰۰-۲۴۱-۱۱۸-۱

تلفن: ۶۶۴۱۵۵۶۱ و ۶۶۴۱۵۵۶۴

همراه: ۰۹۱۲۱۸۸۳۵۸۱

فهرست مطالب

۱۱	پیشگفتار
۱۳	فصل ۱
۱۳	تعاریف و تاریخچه
۱۳	نظام‌های بازیابی اطلاعات: تعاریف
۱۸	تاریخچه نظام بازیابی اطلاعات
۱۸	مقدمه
۱۸	تاریخچه نظام‌های بازیابی اطلاعات
۱۹	دوره اول: قبل از ۱۹۴۰
۲۲	دهه ۱۹۴۰
۲۳	دهه ۱۹۵۰
۲۴	دهه ۱۹۶۰
۲۷	دهه ۱۹۷۰
۳۱	دهه ۱۹۸۰
۳۴	دهه ۱۹۹۰
۳۷	هزاره سوم: سال ۲۰۰۰ تا کنون
۴۱	خلاصه فصل
۴۳	فصل دوم
۴۳	ساختار نظام‌های بازیابی اطلاعات با تأکید بر موتورهای جستجو
۴۳	۱.۲ مقدمه
۴۴	۲.۲ موتور جستجو چیست؟
۴۵	۳.۲ اثربخشی در موتورهای جستجو: یک نیاز اصلی

۴۵	کیفیت نمایه:
۴۷	ب- کیفیت نتایج:
۴۸	ج- کاربردپذیری (قابلیت استفاده):
۴۹	کیفیت ویژگی های جستجو
۴۹	۴.۲ کارآیی موتورهای جستجو
۵۰	۵.۲ معماری موتور جستجو
۵۱	۱.۵.۲ فرآیند نمایه سازی در موتورهای جستجو:
۵۱	۱.۱.۵.۲ مجموعه سازی متنی
۵۳	۲.۱.۵.۲ بخش تغذیه یا فید:
۵۴	۳.۱.۵.۲ تبدیل:
۵۵	۴.۱.۵.۲ مدل داده های سندی:
۵۵	۲.۵.۲ تبدیل متن:
۵۵	۱.۲.۵.۲ نمایه:
۵۶	۲.۲.۵.۲ بازه زمانی رکاز:
۵۷	۳.۲.۵.۲ ریشه گیری
۵۸	۴.۲.۵.۲ تحلیل پیوند
۵۹	الگوریتم رتبه بندی صفحات:
۶۲	۵.۲.۵.۲ استخراج اطلاعات
۶۵	۶.۲.۵.۲ طبقه بندی کننده
۶۶	۳.۵.۲ ایجاد نمایه:
۶۶	۱.۳.۵.۲ آمارهای مدارک:
۶۶	۲.۳.۵.۲ وزن دهی
۶۷	۲.۳.۵.۲ مقلوب سازی:
۶۸	۳.۳.۵.۲ توزیع نمایه
۷۰	۶.۲ فرآیند پرس و جو در موتورهای جستجو
۷۱	۱.۶.۲ تعامل کاربر:
۷۱	۱.۱.۶.۲ درونداد پرس و جو
۷۳	۲.۱.۶.۲ تبدیل پرس و جو
۷۴	۳.۱.۶.۲ برونداد نتایج
۷۴	۲.۶.۲ رتبه بندی
۷۴	۱.۲.۶.۲ نمره دهی:
۷۵	۲.۲.۶.۲ بهینه سازی
۷۵	۳.۲.۶.۲ توزیع
۷۵	۶.۲.۳ ارزیابی:

۷۶	 ۱.۳.۶.۲ گزارش سوابق
۷۶	 ۲.۳.۶.۲ تحلیل رتبه‌بندی
۷۶	 ۳.۳.۶.۲ تحلیل عملکرد
۷۸	 خلاصه فصل:
۷۹	 فصل سوم
۷۹	 مدل‌های بازایی اطلاعات
۷۹	 ۱-۳ مقدمه
۸۱	 ۲-۳ مدل‌های کلاسیک
۸۱	 ۱-۲-۳ مدل بولی
۸۶	 ۲-۱-۲ مدل برداری
۸۹	 ۳-۱-۱-۲ مدل‌های فضا برداری
۹۲	 ۲-۱-۳ مدل احتمالاتی
۹۴	 ۲-۲-۲ مدل‌های مختلف نظریه مجموعه‌ای
۹۵	 ۲-۲-۲-۲ مدل مجموعه بازی
۹۸	 ۳-۲-۲-۲ مقایسه منطق فازی بولی
۹۹	 ۷-۲-۲-۲ مدل بولی گسترش یافته
۱۰۰	 ۳-۲-۳ مدل‌های جبری بدیل
۱۰۰	 ۱-۳-۲ مدل فضا برداری تعمیم یافته (جهتی نام)
۱۰۱	 ۲-۳-۲ مدل نمایه سازی نهان معنا شناختی
۱۰۳	 ۳-۳-۲ مدل شبکه عصبی
۱۰۶	 ۴-۳-۲ مدل‌های احتمالاتی بدیل
۱۰۶	 ۱-۴-۳-۲ شبکه‌های بیزی
۱۰۸	 ۲-۴-۳-۲ مدل شبکه استنباطی
۱۰۹	 ۳-۴-۳-۲ مدل شبکه باوری
۱۱۱	 ۵-۳-۲ مدل‌های بازایی متن ساختارمند:
۱۱۱	 ۱-۵-۳-۲ مدل مبتنی بر سیاهه‌های ناهمبوشان:
۱۱۲	 ۲-۵-۳-۲ گروه‌های ناهمجوار
۱۱۳	 ۶-۳-۲ مدل‌های مرور (تورق):
۱۱۴	 ۱-۶-۳-۲ مدل مرور مسطح
۱۱۴	 ۲-۶-۳-۲ مدل مرور (تورق) ساختار یافته
۱۱۵	 ۳-۶-۳-۲ مدل تورق فرامتن
۱۱۵	 ۷-۳-۲ مدل‌های بازایی بر پایه هستی‌شناسی
۱۱۷	 خلاصه فصل:

۱۱۹	فصل چهارم
۱۱۹	ربط
۱۱۹	۱.۴ مقدمه
۱۲۰	۲.۴ مفهوم ربط
۱۲۲	۳.۴ ابعاد ربط
۱۲۲	۱.۳.۴ بعد اول: منابع اطلاعاتی
۱۲۳	۲.۳.۴ بعد دوم: بازنمایی مشکل کاربر
۱۲۵	۳.۳.۱ بعد سوم: زمان
۱۲۶	۴.۳.۱ بعد چهارم: اجزا
۱۳۰	۵.۴ تضاد ربط
۱۳۲	۶.۴ بازکردن ربط
۱۳۳	۱.۶.۴ بازخورد مثبت
۱۳۴	۲.۶.۴ بازخورد شبه ربط
۱۳۴	خلاصه فصل
۱۳۷	فصل پنجم
۱۳۷	معیارهای ارزیابی
۱۳۷	۱.۵ مقدمه
۱۳۸	۲-۵ سطح و نوع ارزیابی در نظام‌های بازیابی اطلاعات
۱۴۲	۳-۵ مجموعه استاندارد آزمون برای ارزیابی نظام بازیابی اطلاعات
۱۴۲	۱-۳-۵ مجموعه آزمون استاندارد کرافیلد
۱۴۴	۲-۳-۵ مجموعه آزمون استاندارد ترک
۱۴۶	۳-۳-۵ مجموعه آزمون روترز
۱۴۶	۴-۳-۵ مجموعه آزمون کلف
۱۴۷	۵-۳-۵ مجموعه آزمون لیزا
۱۴۷	۶-۳-۵ مجموعه آزمون همشهری
۱۴۹	۴-۵ ربط
۱۵۰	۵-۵ معیارهای ارزیابی نظام‌های بازیابی اطلاعات
۱۵۳	۱-۵-۵ معیارهای سنجش عملکرد
۱۵۳	۱-۱-۵-۵ سنجش عملکرد بازیابی اطلاعات سنتی
۱۵۳	۱-۱-۱-۵-۵ ارزیابی مجموعه‌های بازیابی رتبه‌بندی نشده
۱۵۴	۱-۱-۱-۵-۵ دقت

۱۵۵	۲-۱-۱-۱-۵-۵ بازیافت
۱۵۶	۳-۱-۱-۱-۵-۵ معیار اف
۱۵۸	۵-۱-۱-۱-۵-۵ دقت در n
۱۵۸	۲-۱-۱-۵-۵ ارزیابی نظام‌های بازیابی اطلاعات با نتایج بازیابی رتبه‌بندی شده
۱۵۹	۱-۲-۱-۱-۵-۵ معیارهای ارزیابی ربط دودویی
۱۵۹	۱-۱-۲-۱-۱-۵-۵ متحنی‌های بازیافت - دقت
۱۶۰	۲-۱-۲-۱-۱-۵-۵ دقت درونی
۱۶۱	۲-۱-۲-۱-۱-۵-۵ دقت متوسط میانگین (مب)
۱۶۳	۵-۱-۲-۱-۱-۵-۵ دقت متوسط میانگین هندسی (جی‌مب)
۱۶۴	۶-۱-۲-۱-۱-۵-۵ اولویت دودویی
۱۶۵	۷-۱-۲-۱-۱-۵-۵ دقت در K
۱۶۷	۹-۱-۲-۱-۱-۵-۵ متحنی ROC و حساسیت
۱۶۹	۱۰-۱-۲-۱-۱-۵-۵ میانگین دقت
۱۷۰	۱۱-۱-۲-۱-۱-۵-۵ معیارهای دیگر
۱۷۱	۱-۱-۱-۵-۵ معیارهای ربط چند سطحی
۱۷۳	۱-۲-۱-۱-۱-۵-۵ دقت و بازیافت تعمیم‌یافته
۱۷۳	۲-۱-۱-۱-۵-۵ میانگین رتبه متقابل
۱۷۴	۱۳-۲-۲-۱-۱-۵-۵ بخش
۱۷۴	۱۴-۲-۲-۱-۱-۵-۵ بخش (تعمیم‌یافته)
۱۷۵	۵-۲-۲-۱-۱-۵-۵ دقت پایانه‌ها: کامل، برتر، مفید، واقعی و دقت تفاضلی
۱۷۹	۲-۱-۵-۵ سنجش عملکرد بر اساس اطلاعات تعاملی
۱۸۰	۱-۲-۱-۵-۵ بازیافت تعاملی
۱۸۱	۲-۲-۱-۵-۵ دقت تعاملی ترک
۱۸۱	۳-۲-۱-۵-۵ دقت تعاملی کاربرد
۱۸۱	۴-۲-۱-۵-۵ ربط نسبی
۱۸۲	۵-۲-۱-۵-۵ نشانگر نیمه‌عمر رتبه‌بندی‌شده
۱۸۲	۳-۱-۵-۵ معیارهای رتبه‌بندی و ربط چندسطحی
۱۸۲	۱-۳-۱-۵-۵ طول جستجوی مورد انتظار
۱۸۲	۲-۳-۱-۵-۵ مدت مورد انتظار جستجو
۱۸۳	۳-۳-۱-۵-۵ میانگین طول جستجو
۱۸۳	۴-۳-۱-۵-۵ دقت فوری
۱۸۳	۵-۳-۱-۵-۵ سودمندی (افزایش) جمعی
۱۸۵	۴-۱-۵-۵ معیارهای زمانی
۱۸۵	۱-۴-۱-۵-۵ سرعت جستجو
۱۸۵	۲-۴-۱-۵-۵ سرعت جستجوی واجد شرایط
۱۸۶	۳-۴-۱-۵-۵ زمان پاسخ نظام
۱۸۶	۵-۱-۵-۵ معیار اطلاع‌بخشی
۱۸۶	۶-۱-۵-۵ معیارهای هزینه و سودمندی
۱۸۷	۲-۵-۵ معیارهای زمینه‌ای

- ۱۸۷-۵-۲-۱ ویژگی های کاربر..... ۱۸۷
- ۱۸۷-۵-۲-۲ سنجش نیازهای اطلاعاتی:..... ۱۸۷
- ۱۸۸-۵-۲-۱-۱ سنجش وظیفه مدار:..... ۱۸۸
- ۱۸۹-۵-۲-۲-۲ سنجش موضوع مدار..... ۱۸۹
- ۱۹۰-۵-۲-۲-۳ تداوم و فوریت نیاز اطلاعاتی..... ۱۹۰
- ۱۹۱-۵-۲-۲-۴ مرحله اطلاعاتی..... ۱۹۱
- ۱۹۱-۵-۲-۲-۵ هدف، مقصود و استفاده مورد انتظار از نتایج..... ۱۹۱
- ۱۹۲-۵-۲-۳ معیارهای تعامل..... ۱۹۲
- ۱۹۲-۵-۳ ارزیابی کاربردپذیری..... ۱۹۲
- ۱۹۴-۵-۳-۱ اثربخشی..... ۱۹۴
- ۱۹۴-۵-۳-۲ کارایی:..... ۱۹۴
- ۱۹۴-۵-۳-۳ قابلیت مندی:..... ۱۹۴
- ۱۹۵-۵-۳-۴ زمان تکمیل وظیفه نرمال:..... ۱۹۵
- ۱۹۶-۵-۳-۵ رنخست کاربر نرمال..... ۱۹۶
- ۱۹۶-۵-۳-۶ معیارهای دیگر ارزیابی:..... ۱۹۶
- ۱۹۷-۵-۳-۱-۱ جعبه ابزار..... ۱۹۷
- ۱۹۸-۵-۳-۲-۱ تلاش ذهنی..... ۱۹۸
- ۱۹۹-۵-۳-۲-۲ جریان..... ۱۹۹
- ۲۰۰-۵-۳-۲-۳ تحول شناسی..... ۲۰۰
- ۲۰۰-۵-۶ خلاصه فصل..... ۲۰۰
- ۲۰۱- منابع..... ۲۰۱

پیش گفتار

نزدیک به شش دهه از تولد اصطلاح "بازیابی اطلاعات" می‌گذرد اما هنوز موضوع بحث پژوهش است. جستجوی عبارت "اصول بازیابی اطلاعات" در آمازون^۱ ۳۹۰ عنوان کتاب را در سال ۲۰۱۴ نشان می‌دهد که بیش‌تر آن‌ها در حوزه کتابداری و اطلاع‌رسانی^۲ (۴۶ مورد) است. قدیمی‌ترین کتاب این حوزه در آمازون "مستندسازی و بازیابی اطلاعات: مقدمه‌ای بر اصول پایه و تحلیل هزینه" از پری^۳ و کنت^۴ (۱۹۵۷) است. یکی از اولین و معتبرترین کتاب‌های این حوزه، کتاب "اصول بازیابی اطلاعات" از کوهرن^۵ (۱۹۶۱) است که به توسعه نظریه، محیط نظام بازیابی اطلاعات، اطلاعات و بازنمایی موضوعات، مستندسازی، استفاده و استفاده‌کننده‌ها، راهنماها، و مسائل و موضوع‌های بازیابی اطلاعات پرداخته است.

قدیمی‌ترین کتاب ثبت شده در این زمینه در فهرست کتابخانه کنگره آمریکا همان کتاب پری و کنت (۱۹۵۷) و جدیدترین آن، ویرایش دوم کتاب چاودری^۶ (۲۰۱۰) است (جستجو در سال ۲۰۱۴). یکی از مهم‌ترین کتاب‌های این حوزه، کتاب مفاهیم بازیابی اطلاعات^۷ از پائو^۸ (۱۹۸۹) است که هنوز مرجع بسیاری از متخصصان بازیابی اطلاعات است که آزاد و فتاحی (۱۳۷۸) آن را با همین عنوان ترجمه کرده‌اند.

جستجوی کتاب‌های حوزه بازیابی اطلاعات در کتابشناسی ملی ایران نشان

1 Principle of Information retrieval

2 Amazon: www.amazon.com

3 Documentation and Information Retrieval: An Introduction to Basic Principle and Cost Analysis

4 Perry

5 Kent

6 Kochen

7 Chawdhury

8 Concepts of Information Retrieval

9 Pao

می‌دهد که کتاب‌های فراوانی در این حوزه تالیف یا ترجمه شده است. جدیدترین آن‌ها کتاب‌هایی از شریف‌نژاد، غیاث‌آبادی‌فراهانی و بهشتی (۱۳۹۰)، میدو^۱، بویس^۲، کرفت^۳ و باری^۴ (۲۰۰۷) ترجمه حریری (۱۳۹۰)، یوسفی (۱۳۸۸)، بیزا-ییتس^۵ و ریبیرو-نتو^۶ (۱۹۹۹) ترجمه قاسمی در سال ۱۳۸۵ است که هر کدام ویژگی‌های خاص خود را دارند. کتاب حاضر نیز با توجه به اهمیت موضوع بازیابی اطلاعات تهیه و تدوین شده است. در این کتاب سعی شده است مباحث کلی به زبان ساده اما تخصصی بیان شود تا خوانندگان با اصول کلی نظام‌های بازیابی اطلاعات در عصر - پید آید آشنا شوند.

نمایش شده است تا حد ممکن از تکرار مباحث پرهیز شود و با زبانی ساده پنج موضوع اصلی حوزه بازیابی اطلاعات بررسی گردد. در فصل ۱، تاریخچه و تعاریف نظام‌های بازیابی اطلاعات، در فصل ۲، به طراحی نظام‌های بازیابی اطلاعات توجه شده است. تأکید بر انتخاب موتورهای جستجو از آن جهت بود که موتورهای جستجو یکی از پرکاربردترین ابزارهای بازیابی اطلاعات است. از طرف دیگر، در هر نظام بازیابی نیز موتور جستجو مهم‌ترین بخش است. مبحث فصل ۳، مدل‌های بازیابی اطلاعات است. در این فصل سعی شد مدل‌های بازیابی اطلاعات بر اساس یک نظام مشخص بررسی شود. در فصل ۴، موضوع "رابط" از بیش‌ادترین مفاهیم حوزه بازیابی اطلاعات بررسی شده است. در فصل ۵ سعی شده است معیارهای ارزیابی به شکلی اصولی و نظام‌مند به زبانی ساده بیان شوند.

این اثر می‌تواند در درس‌های مرتبط با بازیابی اطلاعات در رشته‌های مختلف بخصوص علم اطلاعات و دانش‌شناسی استفاده شود. امیدواریم مورد توجه صاحب‌نظران، پژوهشگران و دانشجویان این حوزه قرار گیرد. از بازخوردهای مثبت و منفی آن‌ها در ویرایش‌های بعد استفاده کنیم.

1 Midow

2 Boyce

3 raft

4 arry

5 Ribério- Neto

5 Baeza-Yatez

6 Ribério- Neto