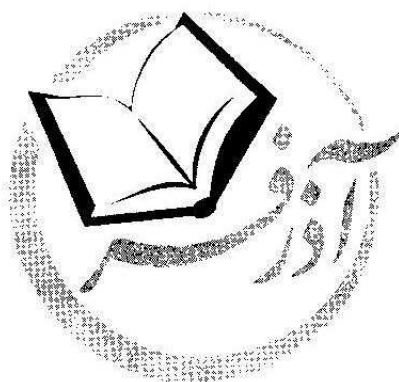


نمایه سازی در وب و موتورهای جستجو

نویسنده:

شیما نوروزبان

مهندس فناوری اطلاعات



سرشناسه: نوروزیان، شیما، ۱۳۷۰.
عنوان و نام پدیدآور: نمایه سازی در وب و موتورهای جستجو/
نویسنده شیما نوروزیان
مشخصات نشر:
مشخصات ظاهری: ۱۶۸ ص.
شابک: ۷-۰۹-۸۹۷۶-۶۰۰-۹۷۸
قیمت: ۱۵۰۰۰۰ ریال
وضعیت فهرست نویسی: فیا
موضوع: رب، موتورهای جستجو
رده بندی کنگره: ۱۳۹۶ ن ۹/ن ۸۸۴/۵۱۰۵ TK
رده بندی دی.بی. ۷۸: ۰۰۱
شماره کتابشناس ملی: ۵۰۳۹۰۱۰

نام کتاب:

نمایه سازی در وب و موتورهای جستجو

نام نویسنده:

شیما نوروزیان

ناشر: آذرفر

طراح و صفحه آرا: حسین افسر

نوبت و سال چاپ: اول، ۱۳۹۶

قطع: وزیری

شمارگان: ۱۰۰۰ نسخه

قیمت: ۱۵۰۰۰۰ ریال

شابک: ۷-۰۹-۸۹۷۶-۶۰۰-۹۷۸

تمام حقوق چاپ و نشر این اثر محفوظ است.

مرکز پخش: شهرری، خیابان قم، پاساژ وکیل، طبقه اول، واحد ۷۵

تلفن مرکز پخش: ۰۹۱۹۲۳۰۰۹۹۲-۵۵۹۲۷۶۰۶

تلفن دفتر مرکزی: ۶۶۱۲۶۵۴۴

فصل اول: نمایه سازی

۱۳	 نمایه
۱۴	 نمایه سازی
۱۴	 نمایه ساز
۱۵	 هدف نمایه ساز
۱۵	 کارکرد نمایه سازی
۱۵	 زبان نمایه سازی
۱۶	 سیستم های نمایه سازی
۱۸	 انواع نمایه سازی
۲۱	 مراحل نمایه سازی
۲۲	 استفاده از اصطلاح نام در نمایه سازی
۲۵	 مقایسه نمایه سازی با چکیده نویسی و فهرست نویسی
۲۷	 نمایه سازی چیست
۲۷	 چگونه نمایه سازی می شود
	فصل دوم: ساختار نمایه سازی در موتورهای جستجو
۳۰	 نمایه سازی وب چیست
۳۱	 نمایه سازی وب
۳۳	 نمایه های وب سایت چگونه ساخته می شوند
۳۷	 معیارهای انتخاب سایت ها برای نمایه سازی
۳۷	 رویکردهای حرفه ای
۳۷	 روش نمایه انتهای کتاب در وب
۳۹	 فراداده ها و نمایه سازی وب
۴۰	 دو عامل مهم در نمایه سازی وب

فصل سوم: ابزارهای نمایه سازی در وب

۴۴	ابر داده
۴۹	شاخه های موضوعی
۵۰	فناوری موتورهای کاوش ساختار متمرکز
۵۲	ساختار پراکنده
۵۳	ریزش کا ب در بانک های اطلاعاتی

فصل چهارم: جستجو

۵۶	راهبردهای جستجو
۵۶	ساختار ابزارهای بازبازی اطلاعات
۵۷	ویژگی های مشترک جستجو
۵۸	انواع جستجو
۷۴	سایر امکانات جستجو
۸۰	راهکارهای بهبود جستجو
۸۱	نکاتی در رابطه با جستجو
۸۱	ابزارهای جستجو
۸۶	ساختار موتورهای جستجو

فصل پنجم: نمایش نمایه

۱۰۸	بازنمود نمایه در وب
-----	---------------------------

فصل ششم: جستجوی اطلاعات فارسی در اینترنت

۱۱۳	خط فارسی
۱۱۴	نقایص عمومی خطوط
۱۱۵	نقایص معمولی خطوط فارسی
۱۲۱	زبان و خط فارسی در اینترنت

۱۲۴	نکات کلیدی جستجو به زبان فارسی
۱۲۷	استانداردهای یونی کد و نویسه های فارسی
	فصل هفتم: وب نامرئی و بازیابی اطلاعات
۱۳۵	بخش های مختلف وب نامرئی
۱۳۸	منابع موجود در وب نامرئی
۱۳۸	اهمیت وب نامرئی
۱۳۹	دلایل عدم بازیابی نمایه سازی وب نامرئی توسط موتورهای جستجو
۱۳۹	شیوه های اطلاع یابی در وب نامرئی
	فصل هشتم: چگونه به وب بخشیدن به نمایه سازی موتورهای جستجو
۱۴۴	ایجاد محتوای منحصر به فرد و بویا
۱۴۵	اضافه کردن نقشه سایت به Google Webmaster Tools
۱۴۵	درخواست از Google برای کاوش کردن سایت شما
۱۴۵	ایجاد یا به روز رسانی پروفایل های اجتماعی
۱۴۶	کمک به عنکبوت گوگل برای کاوش در سایت شما
۱۴۶	حفظ ساختار سایت و ناوبری آسان
۱۴۷	ایجاد محتوای تشکر
	فصل نهم: چالش های موجود بر سر کارایی موتورهای جستجو
۱۵۰	مشکلات نمایه سازی در وب
۱۵۳	رویکرد های جایگزین
۱۵۵	نمونه (چرا صفحات وب سایت من در گوگل نیست؟)
۱۵۹	در نهایت
۱۶۳	منابع

با پیدایش دنیای مجازی شاهد ظهور انقلاب دیگری تحت نام انقلاب اطلاعات هستیم. از سوی دیگر حجم انبوه اطلاعات، کاربران را بر آن داشته تا به کیفیت اطلاعات بیش از کمیت آن بیندیشند. اطلاعات روز آمد، دقیق و معتبر سه ویژگی مهمی است که لازمه ی کار هر محقق می باشد؛ بنابراین رشد فناوری های نو ظهور، روش های جدیدی را به منظور سازماندهی اطلاعات جهت بازیابی دقیق آن به یاری می طلبد.

رشد تصاعی فضای اطلاعاتی وب، آلودگی اطلاعات را دامن زده است. از سوی دیگر وجود اطلاعات میسر و سهیل، اهمیت نمایه سازی وب را بیش از پیش مطرح می سازد؛ بنابراین آنچه که برای کاربران وب مهم است دستیابی سریع و آسان به اطلاعات دقیق، روزآمد و دارای اعتبار می باشد نه حجم انبوه اطلاعات.

با استفاده از روش های نوین سازماندهی اطلاعات مفید از اطلاعات سیال و ناپایدار استخراج شده و کاربر به سهولت می تواند به اطلاعات مورد نیاز خود دسترسی پیدا کند. یکی از مهم ترین این راهکارها، نمایه سازی وب است.

نمایه سازی کارآمد، به کاربر امکان بازیابی اطلاعات دقیق، صحیح و روزآمد را می دهد. باید دانست که نمایه سازی وب مفهوم جدیدی نیست. به بیان دیگر نمایه سازی وب فرایند پالایش و فیلتر اطلاعات دنیای وب است و نمایه ساز پالاینده اطلاعات آن.

دنیای وب با کمک از این راهکار در پیچه ای نو را به روی کاربران می گشاید. در این کتاب می کوشیم تا با نگاهی به نمایه سازی وب، به بررسی انواع نمایه، نمایه سازی در موتورهای جستجو، ابر موتورها و ابر داده ها از جمله ابر داده دوبلین کور، باز نمود نمایه در وب و معرفی نرم افزارهای نمایه سازی و... بپردازیم و بستری مناسب جهت اهمیت و نقش آن در بازیابی اطلاعات جهان مجازی فراهم آوریم.

از زمان پایه گذاری وب جهانی در اواخر دهه ۱۹۸۰ هیچ کس نمی توانست وضعیت و تأثیرات فعلی آن را پیش بینی کند. رونق وب و رشد فزاینده آن بر کسی پوشیده نیست، به نحوی که فقط میزان اطلاعات متنی قابل دسترس آن در حدود ۱ ترابایت تخمین زده می شود. اندازه گیری حجم اطلاعات بر روی شبکه اینترنت به ویژه وب کار بسیار دشواری است. برطبق آمارهایی که گروه Cyveillance عرضه کرده است، بیش از ۱/۲ میلیارد صفحه اطلاعات، بدون تکرار و قابل دسترس (تا اکتبر ۲۰۰۴) بر روی وب موجود بوده و براساس همین مطالعات نرخ رشد انفجاری صفحات وب ۷ میلیون در روز بوده است.

این بدان معناست که در حال حاضر تعداد صفحات وب، به میزان سه برابر آن افزایش یافته است و این اطلاعات براساس آمار (NetCraft 2002) به وسیله بیش از ۲۷ میلیون سرویس دهنده وب در اختیار دسترس قرار می گیرد.

همه دلایل مربوط به چرایی سارانه شدن اطلاعات در محیط چاپی با شدت بیشتری در محیط الکترونیکی صادق است.؛ میلیونها نفر در سراسر جهان نیازهای روزمره خود را از طریق وب مرتفع می کنند.

با گسترش حجم انبوه اطلاعات و به اشتراک گذاری آن و در یک روز افزون برقراری ارتباطات از طریق اینترنت بحث سازمان دادن اطلاعات به قصد سهولت در زیاده مطروح می باشد. حجم وسیع اطلاعات بر روی شبکه وب باعث می گردد تا پاسخ دهی به دارس های ارسالی از سوی کاربران بدون دسترسی به تمام متون و فقط با استفاده از فایل های نمایه صورت گیرد؛ زیرا در غیر این صورت یا باید نسخه ای از اطلاعات درخواستی به صورت محلی ذخیره گردد و یا تمام صفحات از راه دور و از طریق شبکه، در هنگام جستجو دسترس پذیر باشد که این روش ها بسیار گران و کند است.

مسئله اساسی در چنین وضعیتی این است که چگونه می توان بدنه ساختارنیافته و به سرعت رشد کننده وب را تحت کنترل و مدیریت درآورد.

تمام این عوامل، تأثیر، اهمیت و تلاش برای بهبود روش های نمایه سازی و الگوریتم های جستجو را مشخص می سازد.

تاریخچه نمایه سازی (Indexing history)

کهن ترین بایگینه ای که از نمایه سازمان دار در تاریخ سراغ داریم، سیاهه لوحه های گلین کتابخانه های اک و سومر است. اما نخستین گامهای جدی در راه تدوین نمایه در خصوص کتب مقدس دینی رده رده به هزاردهم میلادی در چند صومعه صورت گرفت. در مشرق زمین، خاصه در قلمرو تمدن اسلامی کتبهایی در زمینه های رده بندی، فهرست نویسی و نمایه سازی انجام گردید. فهرست ابن ندیم یکی از این نمونه هاست.

پس از اختراع چاپ، نخستین نمایه نظام مند تدوین شد. «کنراد گستر» فهرست رده بندی شده ای از ۲۰۰۰ کتاب را در سال ۱۵۴۸ منتشر کرد. در قرن هجدهم با پیدایش مفهوم رده بندی، فهرست نویسی و رده بندی یکی از هدفهای مهم علوم گردید. در قرن نوزدهم نیز دانش کتابداری و اطلاع رسانی به شکوفایی رسید. رده بندی دهی دیویی که در سال ۱۸۷۶ منتشر شد؛ ضمن اینکه مهمترین پیشرفت بشر در تاریخ رده بندی است بیشترین تأثیر را هم بر نمایه سازی گذاشت. در این قرن همچنین انتشار نشریات به سرعت رشد یافت، بنابراین دسترسی به مقالات نیازمند نمایه بود، این امر ابتدا با نمایه نشریات ادواری در آمریکا آغاز گشت.

پس از جنگ جهانی دوم، دامنه تحقیقات و پژوهشها روز به روز گسترده تر گردید و این افزایش تدریجی کنترل اطلاعاتی با استفاده از ابزارها و شیوه های قبلی را غیرممکن می ساخت.

در سال ۱۹۶۱ اولین کتاب تئوریک در زمینه نظام‌های بازیابی اطلاعاتی ارائه شد. انگیزه اصلی تألیف این کتاب از جانب «ویکری» رشد سریع انتشارات و بررسی روش‌های نمایه سازی در کتابداری بود. به تدریج بر حجم این تلاش‌ها افزوده شد تا این که در دهه ۱۹۷۰ تحقیقات گسترده‌ای به روی بانک‌های اطلاعاتی و سیستم‌های خودکار پیوسته صورت گرفت. در طول دهه ۱۹۸۰ استفاده از نظام‌های بازیابی اطلاعات در دو جنبه گسترش پیدا کرد.

جنبه اول گسترش نظام‌های تمام‌متن بود (تا قبل از این دهه در بازیابی اطلاعات فقط از نمایه‌ها و چکیده‌ها استفاده می‌شد).

جنبه دوم گسترش نظام‌های پیوسته مورد استفاده غیرمتخصصان بود.

در دهه ۱۹۹۰ حجم منابع ذخیره‌ای اطلاعات رایانه‌ای افزایش یافت، تصاویر به همراه اطلاعات ارائه شده و راه‌نمایش اطلاعات به گونه‌ای متفاوت تعقیب گردید.

تحول مهمی که در این دهه به وقوع پیوست ظهور اینترنت بود. تا قبل از این تحول تعداد ناشران اطلاعاتی بسیار محدود بود ولی در حال حاضر افراد زیادی در سرتاسر جهان اطلاعات مربوط به خود را تولی‌د و دسته‌بندی می‌کنند و از طریق صفحات خانگی، آنها را در اختیار دیگران قرار می‌دهند.

امروزه حجم انبوه اطلاعات به خصوص اطلاعات موجود بر روی شبکه جهانی وب موجب ظهور جنبه‌های جدیدی در بازیابی اطلاعات گردیده است.