

## اطلاع رسانی:

نشریه فنی مرکز اسناد و مدارک علمی (مرکز اطلاعات و مدارک علمی ایران)

دوره هفتم شماره ۲

### نمایه سازی همارا

نوشته: سوزان آرتانندی\*

ترجمه: فیروزان زهادی\*\*

همارایی به مفهوم عام کلمه، یعنی ترکیب دو یا چند واژه برای بوجود آوردن یک رده موضوعی که با دیگر رده های موضوعی که با همان واژه ها اما به طور منفرد نشان داده شده باشند؛ یا حاصل نوع دیگری از ترکیب واژه های مزبور باشند، تفاوت داشته باشد. به عنوان مثال، وقتی واژه های دانشگاه، کتابخانه ها و مدیریت به ترتیبی مناسب همارایی شوند، رده موضوعی که از ترکیب آنها حاصل می شود، چنین است: مدیریت کتابخانه دانشگاهی.

این نوع همارایی که در مثال بالا شرح آن آمد مربوط است به ضرب منطقی بول، و از آن در ساخت زبانه های نمایه برای نظامهای بازیابی پیش هما را بنحو فزاینده ای استفاده شده است.

در این نظامها، واژه های نمایه غالباً از ترکیب دو واژه یا بیشتر جهت توصیف موضوعهای خاص بدست می آید. این واژه های پیش هماراسته، در همان شکل که نمایه ساز به آنها اختصاص داده است و بدون دستکاری اضافی به عنوان واژه های جستجو نیز بکار می روند.

در نظامهای پس همارا، همارایی در مرحله جستجو صورت می پذیرد. بدین معنی که هنگام نمایه سازی، موضوعهای ترکیبی و پیچیده به مفاهیم ساده تر تجزیه می شوند و به هنگام جستجو موضوعهای پیچیده از طریق همارایی مفاهیم ساده بدست می آیند. این روش بخاطر امکان عرضه ترکیبهای متعددی از واژه ها به نظام انعطاف می بخشد.

با استفاده از مثال بالا می توان گفت که در یک نظام پیش همارا از عبارت "مدیریت کتابخانه دانشگاهی" می توان هم به عنوان واژه نمایه و هم واژه جستجو استفاده کرد. در نظام پس همارا، کل مدارک را می باید کاوید تا مشخص شود هر سه واژه "دانشگاه"، "کتابخانه" و "مدیریت" با هم به چه مدارکی اختصاص یافته است.

اگر به همارایی از دید ابزار زبان نمایه نگاه کنیم، باید آن را در طبقه ابزارهایی که مربوط به دقت در بازیابی هستند جای دهیم. هدف از استفاده از ابزارهای دقت در بازیابی این است که احتمال بازیابی مدارک بی ارتباط به موضوع جستجو را کاهش دهد. همارایی با اخص کردن واژه (به عبارت دیگر محدود تر کردن دامنه رده موضوعی) توانائی نظام را برای حذف مدارک بی ارتباط با موضوع می افزاید. اما این خود به قیمت حذف کردن بخشی از مدارک مربوط به موضوع هم تمام می شود.

\*

*Encyclopedia of Library and Information Science. Vol.*

5. PP. 679-682.

\*\* کتابدار حسینیه ارشاد

از نظر سابقه تاریخی معرفی نمایه سازی همارا تا حدود زیادی مدیون "مارتیمرتاب"<sup>۱</sup> است، هر چند قبل از او عده دیگری هم بودند که در نمایه سازی از این شیوه استفاده می کردند، از جمله "باتن"<sup>۲</sup> در انگلستان و "کوردونیه"<sup>۳</sup> در فرانسه. وقتی "تاب" برای اولین بار در سال ۱۹۵۳ از این روش برای تنظیم یک دسته از مدارک ASTIA (مرکز اطلاعات فنی نیروهای مسلح) استفاده کرد واژگان محدودی بکار برد که از اصطلاحات یک کلمه ای به نام "تکواژه" تشکیل شده بود. تک واژه ها از متن خود مدارک گرفته می شدند و هیچ نوع کنترل واژگانی صورت نمی گرفت.

نظامهای نمایه سازی همارای امروزی بتدریج از همین مفهوم اولیه نضج گرفتند. این نظامها، دیگر منحصر بر اصطلاحات تک واژه ای مبتنی نیستند و از واژه هایی که حاصل ترکیب دو کلمه یا بیشتر باشند هم آزادانه استفاده می شود. این کلمه های مرکب را می توان واژه های پیش هماراسته ای دانست که در حین جستجو به روش پس همارا با هم ترکیب می شوند. واژگانی آزاد بتدریج خود را به واژگانهای کنترل شده یا مقید دادند که در اصطلاحنامه ها طبقه بندی شده اند و ساخت زبان نمایه را نشان می دهند. انواع زبانهای نمایه که در نظامهای نمایه سازی همارا بکار می رود، اغلب شامل ارتباطهای پیچیده اصطلاحات است و از این رو جستجوهای تخصصی بغرنج را میسر می کند.

جستجوهای تخصصی بغرنج با ظهور نظامهای کامپیوتری هم میسر شد. در حالیکه نظامهای نمایه سازی همارای دستی و "پیکابو" تا حد زیادی محدود به استفاده از ضرب منطقی هستند و این اساساً حاصل محدودیتهایی است که در روشهای آنها موجود است، در نظامهای کامپیوتری امکان استفاده از ضرب، جمع و تفریق منطقی بول در ساختن عبارات منطقی برای توصیف مجموعه مدارک مورد جستجو وجود دارد. به عنوان مثال، اگر یک مجموعه از مدارک درباره نمایه سازی باشد (مجموعه A) و مجموعه دیگر در مورد مجلات شیمی (مجموعه B)، در این صورت ضرب منطقی این دو (A • B) جستجوی آن دسته از مدارک نظام را که A و B توصیفگر آنهاست یعنی مدارک مربوط به "نمایه سازی مجلات شیمی" را مشخص می کند. تفریق منطقی این دو (A.B) مدارکی را معین می کند که A توصیفگر آنهاست اما B نیست. و این یعنی مدارکی که موضوع آنها نمایه سازی است به استثنای مدارکی که در آنها از نمایه سازی مجلات شیمی سخن می رود. جمع منطقی این دو (که بیانگر شمول است) (A+B) مدارکی را به ما باز می نمایاند که A یا B و یا هر دو توصیفگر آنها هستند، یا به عبارت دیگر مدارکی که به موضوع نمایه سازی یا موضوع مجلات شیمی و یا نمایه سازی مجلات شیمی پرداخته اند. نمودارهای ون<sup>۴</sup> در شکل شماره ۱ آنچه را که در بالا گفته شد نشان می دهد: در سه نمودار مزبور قسمتهای هاشور خورده به ترتیب نشانه A.B (ضرب منطقی)، A.B (تفریق منطقی) و A+B (جمع منطقی می باشند) (۳).

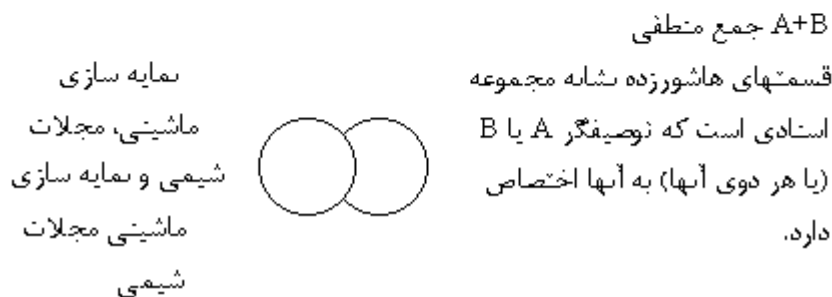
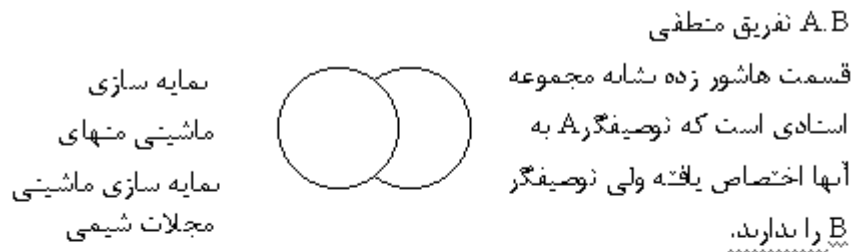
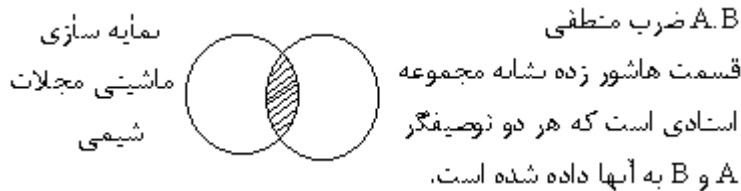
---

<sup>1</sup> - Martimer taube

<sup>2</sup> - Batten

<sup>3</sup> - Cordonnier

<sup>4</sup> - Venn diagrams



بتدریج که نظامهای نمایه سازی همارا به شکل گسترده تر شناخته و بکار گرفته شد بعضی محدودیتها که خاص پس همارایی نمایه ها بود، خود را نشان داد. امکان ترکیب نامحدود واژه ها که تصور می رفت امتیاز این نظام باشد ثابت کرد که در بسیاری از موارد زبان آور است. برای جلوگیری از هماراییهای ناخواسته "رابطها"<sup>5</sup> پیشنهاد شدند و برای نشان دادن عمل یک واژه معین در یک توصیف نمایه ای از "نقشها"<sup>6</sup> استفاده شد. به عنوان مثال، اگر مدرکی هم به موضوع حدیده کاری پلی اتیلن و هم به قالب ریزی پلی پروپیلن پرداخته باشد، در این صورت اگر یک "رابط" بین "پلی اتیلن" و "حدیده کاری" و "رابط" دیگری بین "قالب ریزی" و "پلی پروپیلن" مثل: حدیده کاری  $A$ ، قالب ریزی  $B$ ، پلی پروپیلن  $B$  بگذاریم از همارایی کاذب "حدیده کاری" و "پروپیلن" از یک طرف و "قالب ریزی" و "پلی اتیلن" از طرف دیگر، جلوگیری خواهد کرد. در نتیجه استفاده از رابط، جستجو در مورد حدیده کاری پروپیلن و قالب ریزی پلی اتیلن، منجر به بازیابی کاذب نخواهد شد. عبارت دیگر، ریزش کاذب نخواهیم داشت.

اما، با این که، "رابطها" می توانند از اختلال در امر بازیابی جلوگیری کنند احتمالاً مانع بعضی از هماراییهای بالقوه مفید هم می شوند و این را می توان با مثال کلاسیک و مشهور "تاب" نشان داد و آن مدرکی است که موضوع آن "پوشش سربی لوله های مسی" است. برای اجتناب از همارایی کاذب بین "سرب" و "لوله ها" می توان به ترتیب از رابط بین "سرب" و "پوشش" و نیز "مس" و "لوله ها" استفاده کرد. بهر حال، هر چند این کار احتمال همارایی کاذب را از میان می برد، در عین حال از همارایی بالقوه مفیدی هم جلوگیری می کند و آن

<sup>5</sup> - Links

<sup>6</sup> - Roles

ترکیب دو توصیفگر "پوشش" و "لوله ها" است که پاسخ مناسب جستجوی احتمالی در مورد "پوشش لوله ها" خواهد بود (۴)

هدف از استفاده از "نقش" بالا بردن میزان دقت در نتایج جستجو از طریق بیان روشن ارتباط موجود بین مجموعه های مختلف مدارک است. "نقش" ها با افزودن یک عامل نحوی عمل یک واژه را در توصیف نمایه یا در جستجو مشخص می کنند. مطلب بالا در اصل بدان معنی است که هر واژه معین از طریق شاخصهای متعدد نقش به تعداد زیادی واژه اخص تر تقسیم می شود. تفاوت اصلی این روش با روش استفاده مستقیم از واژه های اخص این است که آن دسته از روابط که "شاخص نقش" بیانگر آنهاست، معمولاً در یک نظام و مجموعه مشخص استاندارد شده و تحت ضابطه قرار می گیرند.

مثلاً، "نظام شورای مشترک مهندسان"<sup>۷</sup> از مجموعه مفصلی از نقشها تشکیل شده است. در این نظام مدرکی که بعنوان مثال در مورد قالب ریزی محفظه های پلاستیکی باشد با اصطلاحات و نقشهایی که در دنبال می آید، نمایه می شود. قالب ریزی (نقش ۸)، مواد پلاستیکی (نقش ۱) و محفظه (نقش ۲). نقش ۸ برای مشخص کردن موضوع اصلی مورد بحث، نقش ۱ برای نشان دادن مواد خام و نقش ۲ برای محصول تولید شده (۵) بکار می رود.

بسیاری از نظامهای نمایه سازی همارا بدون استفاده از "نقش" و "رابط" عمل می کنند. بطور کلی باید گفت، به دلیل مشکلاتی که استفاده یکدست و یکنواخت از "نقش" و "رابط" به بار می آورد، بخاطر آن که کاربرد آنها مستلزم صرف هزینه های اضافی است، چنین می نماید که در نمایه سازی همارا گرایش بسوی عدم استفاده از آنهاست.

با توجه به این امر که استفاده از نقش و رابط باعث افزایش فوق العاده ای در هزینه های نمایه سازی می شود، از نظر اقتصادی، کاربرد آنها فقط هنگامی قابل توجه است که خرجهای اضافی ناشی از کاربرد آنها با کیفیت برتر نتایج بدست آمده جبران شود.

با این حال، استفاده از نقش لزوماً باعث بالاتر رفتن چشمگیر کیفیت بازیابی نمی شود. بررسی اخیر در "سیستم مدلاز"<sup>۸</sup> نشان داد که اگر از شاخص نقش استفاده شده بود کمتر از ۱۰٪ از هماراییهای کاذب قابل جلوگیری بود. برخی نشانه ها نیز حاکی از آن است که کارایی نقش در ارتباط با رشته موضوعی، پیچیدگی جستجو و حجم مدارک تغییر می کند. هر چه زبان رشته موضوعی دقیق تر و ابهام آن کمتر باشد و هر چه جستجو از پیچیدگی بیشتری برخوردار باشد و بالاخره هر چه حجم مدارک زیادتر باشد، بکار گرفتن "نقش" با احتمال موفقیت بیشتری همراه خواهد بود.

احتمال می رود که زبان نمایه راه، که با استفاده از شاخص نقش می توان تخصصی تر کرد، بتوان با اخص تر کردن اصطلاحات نمایه از طرق دیگر هم انجام داد. برای مثال عنوانهای فرعی می توانند هم به شکل نقش و هم به شکل رابط عمل کنند. دست اندرکاران پی برده اند که در نظام مدلاز بسیاری از بی دقتیهای را که در امر بازیابی رخ می دهد و شاخصهای نقش ممکن است تصحیح کنند با اخص تر کردن واژگان می توان از میان برداشت. مثلاً هنگام جستجو در مورد آن دسته از سمهایی که ماهیها تولید می کنند، همارایی واژه های "سمها" و "ماهی" به بازیابی تعدادی مدرک نامربوط به موضوع منجر شد که در آنها سخن از آن دسته از سمهای میکروبی رفته بود که در ماهیها اثر می کند. هر چند ممکن است از این نوع همارایی کاذب با استفاده

---

<sup>7</sup> - Engineers Joint Council System

<sup>8</sup> - Medical Literature Analysis and Retrieval System (MEDLARS)

از نقشه‌های (ماهی به عنوان تولید کننده، ماهی به عنوان موضوع آزمایش) پیشگیری کرد، اما راه دیگری هم برای جلوگیری از آن وجود دارد و آن استفاده از واژه های اخص ترسمه‌های حیوانی و سمه‌های میکروبی است. (۶)

نظامهای پی هم‌را را می توان به طرق مختلف از روش دستی برگه ای گرفته تا روش کامپیوتری، سازمان داد. روش خاص تنظیم مجموعه که در این نوع نظامها بکار گرفته می شود، تا حد بسیار زیادی به روش سازمان دادن آن بستگی دارد. اساس نظامهای نمایه سازی هم‌را هم به شکل دستی و هم به صورت تقارن بصری، معمولاً بایگانی معکوس<sup>۹</sup> است که در آن به هر واژه برگه ای یا محلی برای ثبت آن اختصاص داده می شود. نظامهای کامپیوتری در سالهای اخیر بطور وسیعی بر اساس بایگانیهای مستقیم<sup>۱۰</sup> که روی نوارهای مغناطیسی ثبت شده است عمل می کنند، که در آن، هر محل ثبتي متناظر با یک مدرک است. بعضی از نظامهای عظیم کامپیوتری، ترکیبی از بایگانی مستقیم و معکوس را بکار می برند. بایگانی معکوس به منظور محدود کردن دامنه موضوع، مورد استفاده قرار می‌گیرد و بایگانی مرتب که خود با دامنه محدود در ارتباط است، بکار جستجو مفصل درباره موضوع می آید (۷ و ۸). در آینده افزایش حجم نظامهای نمایه سازی و پیشرفتهایی که در تکنولوژی کامپیوتر صورت خواهد گرفت بدون شک بنحو چشمگیری بر روشهای تنظیم بایگانی نظامهای نمایه سازی هم‌را تاثیر خواهد گذاشت.

[ابتدای صفحه](#)

---

<sup>۹</sup> - Inverted File

<sup>۱۰</sup> - Direct Files