چگونه با چند راهکار ساده بازیابی اطلاعات خود را در پایگاه های اطلاعاتی افزایش دهیم

هر کنش شما در قالب «عبارت جستجو[۱] »، در دنیای اینترنت با واکنشی در قالب « نتایج جستجو» به شما برگردانده خواهد شد. «عبارت جستجو» بیانگر نیاز اطلاعاتی شماست که اغلب در قالب کلمات به نظام ارائه می‌گردد. اگر شما شناختی از نظام‌های بازیابی اطلاعات داشته باشید به شما کمک خواهد کرد تا «عبارت جستجو» خود را با زبان نظام تطبیق دهید؛ اما شناخت نظام بخشی از نگارش « عبارت جستجو» است، بخش دیگر، گزینش واژگانی است که به تقویت این « عبارت جستجو» کمک خواهد کرد. این واژگان از کجا باید انتخاب شوند؟ آیا تاکنون به واژگانی که برای جستجو انتخاب کرده‌ایم فکر کرده‌ایم؟ یادگیری این مباحث در مدیریت زمان و بازیابی اطلاعات مناسب در فرایند جستجو به شما کمک خواهد کرد.

۱-نظام‌های بازیابی اطلاعات از چه روشی برای بازیابی اطلاعات استفاده می‌کنند؟

نظام‌های بازیابی اطلاعات از دو روش کلی برای بازیابی اطلاعات استفاده می‌کنند:

۱-۱. روش بازیابی مبتنی بر انطباق مطلق

دقیقاً «عبارت جستجویی» که شما به نظام ارائه داده‌اید را با کلمات و عبارات مدارک مطابقت می‌دهند و اگر در مدرکی، «عبارت جستجوی» شما تکرار شده بود را برای شما بازیابی می‌کنند. در آغاز نظر بر این بود که وقتی در مدرکی واژه‌ای زیاد تکرار شده باشد احتمال دارد که این مدرک در رابطه با آن موضوع مرتبط‌تر باشد، بنابراین نظام، اول مدارکی را برای شما بازیابی می‌کرد که تعداد تکرار «عبارت جستجوی» شما در آن بیشتر باشد. با گذشت زمان برخی از نویسنده‌ها متوجه این رتبه‌بندی نظام‌ها شدند و سعی می‌کردند برای دیده شدن مطالبشان میزان تکرار واژگان بخصوص را افزایش دهند.

با افزایش نارضایتی کاربران از بازیابی اطلاعات غیر مرتبط، متخصصان بازیابی اطلاعات متوجه شدند که صرف تکرار واژه در یک مدرک نمی‌تواند دلیل معتبری برای مرتبط بودن مدرک با «عبارت جستجو» باشد، بنابراین تصمیم گرفتند که از طریق وزن دهی به واژگان، مدارک را رتبه‌بندی نمایند. روش کار به این صورت بود که واژه در کجای مدرک ذکرشده باشد احتمال دارد با نیاز کاربر همخوانی بیشتری داشته باشد.

مثلاً؛ زمانی که یک کلیدواژه در عنوان ذکر می‌شود چه مقداری می‌گیرد اگر همان واژه در متن، نتیجه‌گیری و یا قسمت‌های دیگر مدرک ذکر شده باشد چه مقداری می‌گیرد. در کل مجموع مقادیری که برای ارزش‌گذاری یک مدرک در رابطه با یک کلیدواژه وجود داشت وزن مدارک را تعیین می‌کرد و هر چه وزن یک مدرک بالاتر بود میزان مطابقت آن با نیاز اطلاعاتی کاربر بالاتر می‌رفت و در اولویت نتایج جستجو قرار می‌گرفت. چیزی که در همه مراحل گذشته شاهد آن هستیم این است که نظام بازیابی اطلاعات مطابق با «عبارت جستجو» عبارت موجود در متن مدارک را مورد ارزیابی قرار می‌دهد. مدل‌های بولی،مدل برداری و مدل احتمالی ازجمله مدل‌هایی بودند که با این رویکرد شکل گرفتند.

۱-۲.روش بازیابی مبتنی بر انطباق نسبی انطباق برتر

با گذشت زمان پژوهشگران متوجه شدند که برخی مدارک هستند که کلیدواژه‌ی خاصی در آن‌ها فقط یک‌بار تکرار شده یا اصلاً عبارت جستجوی کاربر دقیقاً در مدرکی ذکر نشده است اما یک واژه مترادف، مشابه یا مرتبط توانسته است پاسخگوی نیاز اطلاعاتی کاربر باشد؛ بنابراین شیوه‌های جدیدی در کنار شیوه‌های قدیمی به وجود آمد که می‌توانست بر پایه ربط مدارک نتایج جستجو را برای کاربر اولویت‌بندی نماید. مدل‌های مدل فازی، نمایه‌سازی معنایی نهفته، شبکه‌های استنتاجی و الگوهای فرامتن نمونه‌هایی از مدل‌هایی هستند که مطابق با رویکرد جدید شکل گرفتند.

۲- چگونه «عبارت جستجو» را بنویسیم که در کمترین زمان، بیشترین اطلاعات مرتبط را بازیابی کنیم؟

سه‌گام ساده در جستجوی موفق عبارت‌اند از: آماده کردن[۲]، سازمان دادن[۳] و ترکیب کردن[۴]

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

آماده کردن

قبل از شروع یک جستجو، زمانی را برای « تفکر» در رابطه با اطلاعاتی که به دنبالش هستید سپری کنید. هر واژه‌ای که به ذهنتان می‌رسد اعم از اسامی، سازمان‌ها و عبارات را «یادداشت» کنید. در رابطه با کلماتی که به ذهنتان خطور کرده است، در سایت‌های مختلف جستجوی ساده‌ای انجام دهید. برای شروع بهتر است، فقط عناوین، چکیده و کلیدواژه‌های مطالب را مرور کنید. این کار به شما اجازه می‌دهد که واژگان اولیه خود را اصلاح کنید، واژگان نامتناسب را حذف کنید و دامنه جستجوی خود را متناسب با نیاز اطلاعاتی خود محدود کرده یا گسترش دهید.

بعدازاینکه فهرستی از واژگان مرتبط با نیاز اطلاعاتی خود را یافتید برای بهتر شدن نتیجه جستجو به اصطلاح‌نامه‌ها یا سر عنوان‌های موضوعی مراجعه کنید و واژگان دیگری که در آن حوزه اطلاعاتی مرتبط با نیاز اطلاعاتی شماست را انتخاب نمایید و به دایره واژگانی خود اضافه کنید. در بسیاری موارد واژه‌نامه‌هایی که مترادف‌ها را ارائه می‌نمایند نیز می‌تواند مفید باشند.

شاید بپرسید: چرا باید به چنین منابعی رجوع کنیم؟

به خاطر اینکه این احتمال وجود دارد که دو متخصص در یک‌رشته علمی برای بیان یک مفهوم از دو اصطلاح متفاوت استفاده کرده باشند. همچنین احتمال اینکه در یک مفهوم عام‌تر، یا خاص‌تر در رابطه با موضوع شما مطالبی آورده شده باشد وجود دارد. به‌عنوان‌مثال احتمال دارد متخصصی از واژه«تومور» و متخصص دیگری از واژه «سرطان» استفاده کرده باشد و زمانی که شما از هر دو این واژگان استفاده می‌کنید، نتایج جستجو متفاوت خواهد بود. همچنین احتمال دارد در یک مقاله‌ای که در زمینهی خاص‌تری در رابطه با «تومور مغزی» است اطلاعات ارزشمندی در خصوص مفهوم «سرطان» پیدا کنید. همچنین اگر واژه «جهش سلولی» را استفاده کنید نتایج بازیابی متفاوت خواهد بود.

پیشنهاد می‌شود برای دسترسی به سر عنوان‌های فارسی و لاتین به سایت‌های زیر مراجعه نمایید:

http://id.loc.gov/authorities/subjects.html

https://www.ncbi.nlm.nih.gov/mesh

سازمان دادن

برای ساماندهی «عبارت جستجو» ابتدا فهرستی از واژگان حیاتی که متناسب با نیاز اطلاعاتی شما است را ایجاد نمایید، موضوع خود را به مفاهیم جزئی‌تر تقسیم کنید و روابط بین واژگان را تعیین کنید: واژگان عام، خاص و مرتبط

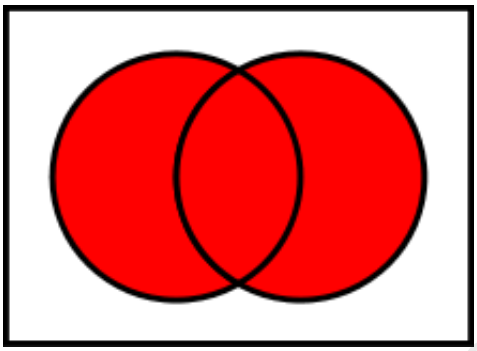
ترکیب کردن

یکی از راه‌های ترکیب کردن کلمات استفاده از عملگرهای بولی است. سه عملگر اصلی بولی عبارت‌اند از: AND، OR و NOT

در لیست واژگان حیاتی شما کلمات مترادف یا مشابه ای هستند که می‌خواهید همه آن‌ها را نظام اطلاعاتی بازیابی کند، برای اتفاق افتادن چنین حالتی شما باید از عملگر OR استفاده کنید. استفاده از این عملگر به این معناست که شما از نظام اطلاعاتی درخواست کرده‌اید که مدارکی را برای شما بازیابی کند که در آن یکی از این واژگان وجود دارد، به عنوان مثال شما با این ترکیب «ورزش» OR «تربیت بدنی» از نظام خواسته‌اید مدارکی را بازیابی کند که یا ورزش و یا تربیت بدنی در آن‌ها ذکرشده است.

«ورزش» OR «تربیت بدنی»

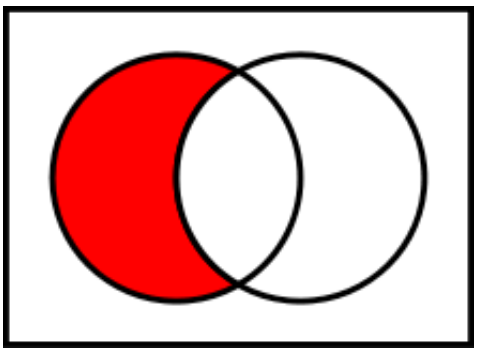
تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.



استفاده از این عملگر مناسب زمانی است که شما بخواهید میزان مدارک بازیابی شده افزایش دهید و به عبارت بهتر اطلاعات موردنیاز شما در یک موضوع محدود است.

چنانچه شما در نتایج بازیابی شده، شاهد مدارکی هستید که نامرتبط است، می‌توانید کلیدواژه مشترک در مدرک نامرتبط را با استفاده از عملگر NOT از جستجوی خود حذف نمایید. به‌عنوان‌مثال شما در جستجوی «ورزش» OR «تربیت‌بدنی» شاهد تعداد زیادی از مدارک هستید که در آن «مدیریت ورزشی" تکرار شده است. برای پاک کردن این واژه از فرمول زیر استفاده کنید:

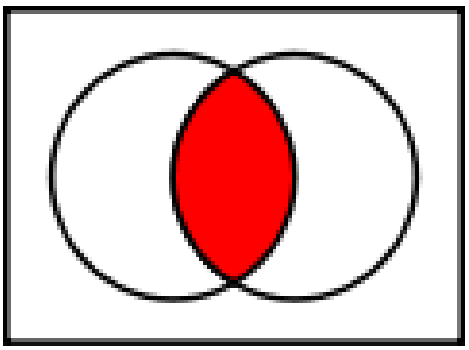
«مدیریت ورزشی» NOT «تربیت‌بدنی»



استفاده از این عملگر مناسب زمانی است که شما بخواهید میزان مدارک بازیابی شده را کاهش داده و مدارک متناسب با نیاز اطلاعاتی خود را انتخاب کنید.

چنانچه شما در جستجوی خود به دنبال واژگانی هستید که هر دو واژه در یک مدرک ذکر شده باشند می‌توانید از عملگر واژه این مستقیم و هستید زنان ورزشی آسیب با رابطه در اطلاعاتی دنبال به شما اگر مثال عنوان به .کنید استفاده AND را جستجو کنید احتمال اینکه تعداد زیادی مدارک برای شما بازیابی شود که در آن واژه آسیب، ورزش و زنان به‌تنهایی به کار رفته است وجود دارد. برای نوشتن این «عبارت جستجو» از فرمول زیر استفاده کنید.

«آسیب ورزشی» AND « زنان »



تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

استفاده از این عملگر مناسب زمانی است که شما بخواهید میزان مدارک بازیابی شده را کاهش داده و مدارک متناسب با نیاز اطلاعاتی خود را انتخاب کنید.

چنانچه تعداد واژگان شما در یک‌طرف یا دو طرف عملگر بیشتر بود از علامت پرانتز برای واژگان قبل و بعد از پرانتز استفاده کنید:

نمونه: (X3 AND X2 AND X1) NOT (Y2 AND Y1)

(«ورزش») OR («مدیریت ورزشی») NOT («تربیت‌بدنی»)

عملگرها باید به صورت درشت‌نمایی و با فاصله نوشته شوند. این سه عملگر قابلیت تبدیل شدن به فارسی را ندارند یعنی نمی‌توانید به جای AND از «و» استفاده کنید.

۳- علائم دیگر برای بازیابی مناسب اطلاعات در محیط وب

روش‌های دیگری برای جستجوی راحت‌تر در محیط وب وجود دارند که می‌توانید با دقت بیشتر و زمان کمتر اطلاعات مورد نیاز خود را بازیابی کنید:

1. با قرار دادن عبارت جستجو در داخل علائم

- علامت " ": زمانی از این علامت استفاده می‌شود که شما به دنبال یک عنوان، عبارت و... هستید که می‌خواهید موتور جستجو دقیقاً آن را برای شما بازیابی نماید نه مشابه آن را.
- علامت{} و []: برای جستجو عبارت به صورتی که زوج‌های کلمات در فاصله نزدیک به یکدیگر قرار داشته باشند.
- علامت– :نادیده گرفتن کلمات (برخی واژه‌ها)
- علامت + : لحاظ کردن کلمات (تمامی واژه‌ها)
- علامت ~ : کلمات متشابه و مترادف
- علامت * : برای یافتن سایر شکل‌های یک واژه

2. اضافه کردن علامت نوع فایل[۵]

به عنوان مثال شما به دنبال پاورپوینت یا پی‌دی‌اف یک موضوع هستید . در این حالت شما بعد از کلمات جستجو نقطه گذاشته و نوع فایل را به انگلیسی اضافه نمایید. به‌عنوان‌مثال: اگر نیاز به پاورپوینت «ورزش زنان» دارید به این طریق عمل کنید

مثال: ورزش زنان.ppt

3. استفاده از کلمهSite : زمانی که به دنبال جستجوی اطلاعات مشخصی از یک سایت خاص هستید می‌توانید بعد از «عبارت جستجو» کلمه Site و سپس آدرس آن را قرار دهید.
4. علامت سه‌نقطه: برای نمایش بازه عددی(سال و قیمت) استفاده می‌شود.

مثال: ورزش زنان ۱۳۸۵...۱۳۹۱ یا مثل TV Sony $300...$500

5. استفاده از واژه define زمانی که می‌خواهید تعریف یک واژه یا عبارت را بازیابی نمایید.
6. علائم مربوط به اعمال اصلی محاسبات ماشین‌حساب
7. استفاده از واژه time: زمان قبل از نام شهر
8. استفاده از واژه weather: آب‌وهوا قبل از نام شهر
9. استفاده از واژه sunrise یا sunset: طلوع و غروب خورشید قبل از نام شهر
10. استفاده از واژه intitle و allintitle: زمانی که می‌خواهید موتور جستجو کلمه دقیقی را در عنوان یک مدرک برای شما بازیابی کند.

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

11. شما برای (سایت یک نشانی) URL در را دقیقی کلمه جستجو موتور خواهیدمی که زمانی :allinurl یا Inurl بازیابی کند.

به‌طورکلی اینکه ما زبان یک برنامه یا نظام را بدانیم کمک می‌کند که ما راحت‌تر با آن نظام ارتباط برقرار کنیم؛ اما ادامه روند ارتباط نیازمند شناخت رفتار و اصلاح آن درروند تعامل است. وقتی «عبارت جستجوی» شما مبهم باشد و دامنه آن به خوبی تعریف نشده باشد، نظام تلاش خود را برای فهم نیاز اطلاعاتی شما به کار خواهد گرفت اما نتایج ممکن است رضایت‌بخش نباشد. برای داشتن یک رابطه سودمند و گرفتن نتایج بهتر از یک نظام اطلاعاتی بهتر است زمان بیشتری را برای نوشتن یک «عبارت جستجو» سپری کرد.

پانویس ها:

Search quary [۱]

Prepare [۲]

Organize [۳]

Combine [۴]

filetype [۵]

منبع: https://nlp.stanford.edu/IR-book/newslides.html

نویسنده: افروز همراهی

مشخصات استناددهی به این مقاله

نویسنده(ها):	افروز همراهی
عنوان مقاله:	چگونه با چند راهکار ساده بازیابی اطلاعات خود را در پایگاه های اطلاعاتی افزایش دهیم
عنوان مجله:	کتابدار ۲.۰ –(عنوان لاتین: i-Kitابdار 2.0)
دوره مجله(Vol):	۶
شماره مجله(Issue):	۲
سال(Year):	۱۳۹۹
شناسه دیجیتال(DOI):	
لینک کوتاه:	http://lib2mag.ir/11463

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.