



ارزیابی جامعیت و مانعیت نظامهای بازیابی اطلاعات

علی اصغر شیرینی

مفاهیم و فرمولها

اصطلاحات "جامعیت" و "مانعیت" به مقیاسهایی اطلاق می‌گردد که عموماً جهت ارزیابی عملکرد کاوش - بخصوص در نوعی نظام بازیابی اطلاعات - به کار می‌رود. درجه جامعیت و مانعیت را با نسبت محاسبه می‌کنند. جامعیت: توانایی نظام در بازیابی مدارک مرتبط است که به صورت ذیل محاسبه می‌گردد.

$$\text{جامعیت} = \frac{\text{تعداد مدارک مرتبط بازیابی شده}}{\text{کل تعداد مدارک مرتبط موجود در مجموعه}} \times 100$$

مانعیت: توانایی نظام در کنار گذاردن مدارک نامرتبط.

$$\text{مانعیت} = \frac{\text{تعداد مدارک نامرتبط بازیابی شده}}{\text{کل تعداد مدارک بازیابی شده}} \times 100$$

هر دو نسبت جامعیت و مانعیت برای بیان قابلیت سیستم در تصفیه اطلاعات کاربرد دارد. بدان معنا که توانایی سیستم در بازیابی آنچه می‌خواهیم و کنار گذاردن آنچه با نیاز ما ارتباطی ندارد، سنجیده می‌شود.

این مقیاسها به روشنی در جدول ۲×۲ ذیل آمده است.

جدول ۱. مقیاسهای سنجش عملکرد بازیابی

تعداد مدارک	مرتبط	نامرتبط	کل
بازیابی شده	a	b	a+b
بازیابی نشده	c	d	c+d
کل	a+c	b+d	a+b+c+d

در اغلب نظامهای بازیابی اطلاعات هنگامی که کاوش صورت می‌گیرد، نظام، مجموعه مدارک را به دو بخش تقسیم می‌کند. مدارکی که با استراتژی کاوش به کار رفته برای جستجو در نظام منطبق است که بازیابی شده اند (a+b) و کلیه مدارکی که با این استراتژی همسان نشده اند و در نتیجه بازیابی نیز نشده‌اند (c+d). این تقسیم بندی دوگانه مجموعه مدارک را می‌توان به عنوان شکلی از پیش بینی تناسب نظام دانست. نظام به تعبیری، پیش بینی می‌کند که مدارکی مرتبط است و مدارک دیگر نامرتبط. بدین ترتیب مدارک مرتبط را بازیابی می‌کند و مدارک نامرتبط را کنار می‌گذارد.

تقریباً در تمام موارد تعداد مدارک بازیابی شده در یک کاوش از کل تعداد مدارک کمتر است. به بیان دیگر تقریباً در تمام کاوشها a+b کاملاً کمتر از c+d است و تعداد مدارک بازیابی نشده بسیار است. در یک کاوش برای مثال ۸۰ مدرک از مجموعه ۵۰۰/۱۰۰۰ مدرکی بازیابی می‌شود، در این حالت

$$a+b = 80 \text{ در حالیکه } c+d = 499/920 \text{ است}$$

جنبه دیگر جدول مذکور به تصمیم گیریهای مربوط به تناسب مدرک با نیاز استفاده کننده نظام باز می‌گردد. یک کاوش کامل، کلیه مدارک موجود در پایگاه اطلاعاتی را که استفاده کننده آنها را مرتبط ارزیابی می‌کند، باز می‌یابد (a+c) و مدارکی را که او مرتبط ارزیابی نمی‌کند بازیابی نمی‌شود (b+d). در این وضعیت همسانی کاملی میان ارزیابی های

استفاده کننده در مورد مدارک مرتبط و پیش بینی های تناسب نظام حاصل می گردد. بدین ترتیب $b = 0$ و $c = 0$ است و می توان گفت که این کاوش ۱۰۰ درصد جامعیت و مانعیت را کسب کرده است.

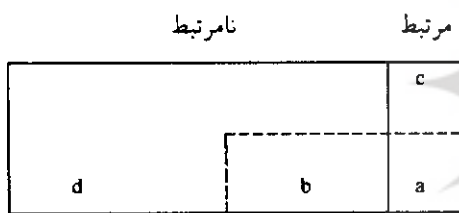
اگر بخواهیم جامعیت و مانعیت را در قالب جدول فوق ببریم بدین صورت خواهد شد:

$$\text{نسبت جامعیت} = \frac{a}{a+c}$$

$$\text{نسبت مانعیت} = \frac{a}{a+b}$$

نسبت مانعیت را می توان نوعی عامل هزینه بر شمرده، چون وقت استفاده کننده به عنوان یک عامل هزینه به شمار می آید، و آن وقتی است که برای تفکیک مدارک مرتبط از نامرتب در یک کاوش صرف می گردد. به عنوان مثال، برای کاوش در یک پایگاه اطلاعاتی ۲۰ مدرک مرتبط وجود دارد. فرض کنیم استفاده کننده ای از سه استراتژی متفاوت کاوش در نظام استفاده می کند و از طریق هر استراتژی ۱۵ مدرک از ۲۰ مدرک را بازیابی می کند (جامعیت ۷۵٪) در کاوش اول، کل تعداد مدارک بازیابی شده ۳۰ عنوان، در استراتژی دوم ۶۰ و در استراتژی سوم ۱۵۰ مدرک است. نسبت مانعیت در این سه کاوش به ترتیب ۵۰٪، ۲۵٪ و ۱۰٪ است. در کاوش اول استفاده کننده باید به بررسی فقط ۳۰ مدرک برای یافتن ۱۵ مدرک مربوط پردازد و در کاوش دوم ۶۰ مدرک و در کاوش سوم ۱۵۰ مدرک را باید بررسی نماید. در شرایط مساوی وقت بیشتری از او صرف می گردد تا مدارک مرتبط را از مدارک نامرتب در جستجوی دوم نماید و همین مسئله در کاوش سوم نیز از کاوش دوم مشهودتر است. بدین دلیل است که نسبت مانعیت به عنوان مقیاس تلاش یا هزینه استفاده کننده به شمار می آید. کاوش که ۷۵٪ جامعیت و ۵۰٪ مانعیت داشته باشد به مراتب مؤثرتر از جامعیت ۷۵٪ و مانعیت ۲۵٪ است و این نیز به مراتب مؤثرتر از جامعیت ۷۵٪ و مانعیت ۱۰٪.

همانطور که در پیش اشاره شد هر دو نسبت جامعیت و مانعیت، میزان انطباق میان ارزیابیهای کاربر از تناسب مدرک و پیش بینی تناسب نظام را مشخص می سازد. در یک کاوش کامل، این دو کاملاً بر هم منطبق می گردد. متأسفانه، چنین کاوشهای دقیقی به ندرت رخ می دهد. ما بیشتر برآنیم که به موقعیتی برسیم که انطباق مقطعی میان مجموعه $a+c$ و مجموعه $a+b$ به وجود آوریم. در نمودار زیر کل مجموعه به صورت مستطیل پیوسته و مدارک بازیابی شده به صورت مستطیل خط چین نشان داده شده است. در یک کاوش فرضی اما کاملاً نمونه، اکثر مدارک مرتبط - نه همد - بازیابی و اکثر مدارک نامرتب - نه همد - کنارگذاشته شده است (۴، ص ۱۷۰-۱۷۳)



نمودار ۱. نتایج یک کاوش نمونه در نظام بازیابی

افزون بر دو مقیاس جامعیت و مانعیت که به عنوان مقیاسهای سنجش عملکرد نظامهای بازیابی اطلاعات به شمار می آیند در برخی متون به مقیاسهای دیگری اشاره شده که ذکر آنها خالی از فایده نیست.

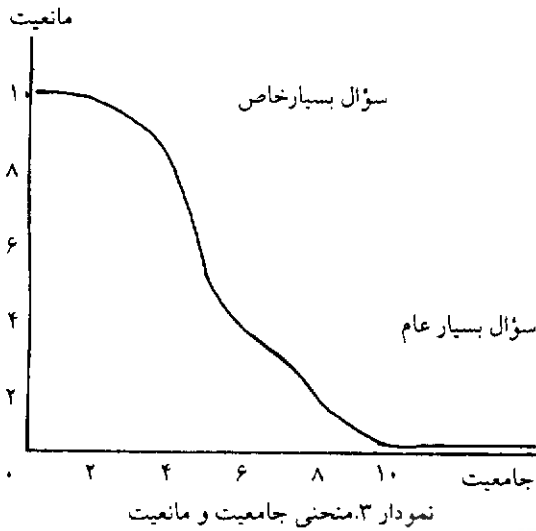
۱. آفت (Fallout). عبارت است از احتمال اینکه یک مدرک بازیابی شده نامرتب باشد، که اگر بخواهیم بر اساس جدول شماره ۱ نشان دهیم بدین ترتیب خواهد بود:

$$\frac{\text{مدارک بازیابی شده نامرتب}}{\text{کل مدارک نامرتب}} \quad \text{یا} \quad \frac{b}{b+d} = \text{رخداد}$$

۲. شمول (Genesality). به معنای نسبت مجموعه مدارک (پایگاه اطلاعاتی) که پرسش بدان می پردازد. فرمول بر اساس جدول ۱:



ارزیابی جامعیت و مانعیت نظام‌های بازیابی اطلاعات



سنجش جامعیت و مانعیت

سنجش مانعیت برای یک کاوش به سهولت انجام می‌گیرد اما اندازه‌گیری جامعیت به سادگی مقدور نیست. یک جدول ۲×۲ دیگری را در نظر بگیرید. این بار جدول نتایج فرضی کامل شده است. برخی از نتایج در جدول مستقیماً قابل مشاهده‌اند و برخی دیگر از نتایج از طریق همکاری درخواست‌کننده قابل حصول است و بعضی تنها قابل برآورد است.

جدول ۲. نتایج یک کاوش فرضی در یک نظام بازیابی

ارزیابیهای استفاده‌کننده در مورد ارتباط مدرک با درخواست			
کل	نامرتبط	مرتبط	
۵۰	۴۰	۱۰	بازیابی شده
۴۹۹۹۵۰	۴۹۹۹۴۵	۵	بازیابی نشده
۵۰۰۰۰۰	۴۹۹۹۸۵	۱۵	جمع کل

اطلاعات

پایگاه اطلاعاتی که در جدول فوق نشان داده شده است دارای ۵۰۰۰۰۰ مدرک است و در کاوش فرضی که صورت گرفته است، ۵۰ مدرک بازیابی شده و ۹۵۰ و ۴۹۹ مدرک

$$\text{کل مدارک موجود در پایگاه اطلاعاتی} = \frac{a+c}{a+b+c+d} \text{ یا } \frac{a+c}{\text{شمول}}$$

۳. اختلال (Noise): عبارت است از نسبت مدارک بازیابی شده نامرتبط (۵، ص ۱۶۵):

$$\text{اختلال} = \frac{b}{a+b} \text{ یا } \frac{\text{تعداد مدارک بازیابی شده نامرتبط}}{\text{کل تعداد مدارک بازیابی شده}}$$

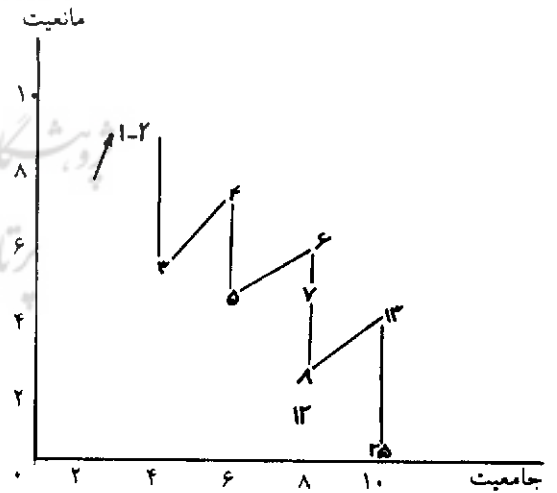
رابطه جامعیت و مانعیت

اساساً دو مقیاس جامعیت و مانعیت در روند کاوش نتیجه معکوس دارند. بدین معنا که هر تلاش در جهت افزایش جامعیت صورت گیرد ناگزیر باعث کاهش مانعیت می‌گردد. هر قدر سعی کنیم که مانعیت را به سطح مطلوبی برسانیم جامعیت تاثیر پذیرفته و کاهش می‌یابد.

عملکرد مطلوب یک سیستم عبارت است از:

$$1 = \text{مانعیت} = \text{جامعیت}$$

یعنی بازیابی کلیه مدارک مرتبط بدون وجود مدارک نامرتبط



نمودار ۲. جامعیت و مانعیت در یک کاوش

از جهت دیگر منحنی جامعیت و مانعیت می‌تواند نمایشگر عملکرد سیستم با توجه به ماهیت پرسش کاوش باشد. سؤال بسیار خاص عملکرد را در بالای منحنی و سؤال بسیار عام عملکرد را در پایین منحنی نمایش می‌دهد.

(ص ۱۶۸-۱۶۹)

از مجموعه "الف" انجام دهیم. زیر مجموعه "الف" شامل تعداد مدارکی است که می‌دانیم با تقاضا مرتبط و در پایگاه اطلاعاتی موجود است.

این زیر مجموعه می‌توانست شامل مدارک مرتبطی باشد که توسط پرسشگر قبل از کاوش مرتبط شناخته شده است، یا شامل مدارکی باشد که از طریق کاوش موازی در یک یا چند نمایه صورت گرفته است. (برای مثال نمایه های چاپی) و توسط پرسشگر مرتبط شناخته شده است، یا می‌توانست با ترکیبی از این دو منبع حاصل گردد. به هر صورت، به این نتیجه می‌رسیم که نسبت جامعیت کاوش بر مبنای تعداد مدارک زیرمجموعه "الف" صورت گرفته که بازتابی می‌شود، به عنوان مثالی ساده، فرض کنیم زیر مجموعه "الف" شامل سه مدرک مرتبط و موجود در مجموعه است. در نتیجه کاوش دو مدرک از سه مدرک بازتابی می‌شود و بنابراین جامعیت کاوش ما ۶۶٪ خواهد بود. در مورد یک کاوش فرضی (جدول ۲) ۱۰ مدرک مرتبط بازتابی شده است. برآورد جامعیت نشان می‌دهد که ۱۰ مدرک حدوداً ۶۶٪ از کل مدارک مرتبط در پایگاه اطلاعاتی است. بنابراین برآورد ما از تعداد مدارک مرتبط، ۵ مدرک مرتبط را شناخته است.

زیر مجموعه "الف" مبنای جامعیت است که از طریق آن می‌توان نسبت جامعیت یک کاوش را برآورد نمود. روشن است که با میناهای بسیار کوچک اطمینان آماری کمتری در مورد برآورد جامعیت یک کاوش خواهیم داشت (۴، ص ۱۷۶ - ۱۷۷).

عوامل موثر بر جامعیت و مانعیت

کل نگری و جزء نگری در نمایه سازی دو عامل اساسی هستند که بر مقیاسهای مانعیت و جامعیت اثر می‌گذارند.

الف. کل نگری (Exhaustivity)

عبارت است از شناسایی کلیه موضوعات مجزای مورد بحث مدرک در عملیات نمایه سازی. برای مثال اگر ما مدرکی در ارتباط با شش موضوع (الف، ب، ج، د، ه، و) داشته و کلیه شش موضوع را در مرحله تحلیل مفهومی نمایه سازی در نظر بگیریم و آنها را به وسیله ترکیب مختلف واژه های نمایه بیان نماییم می‌توانیم بگوییم در نمایه سازی این مدرک خاص

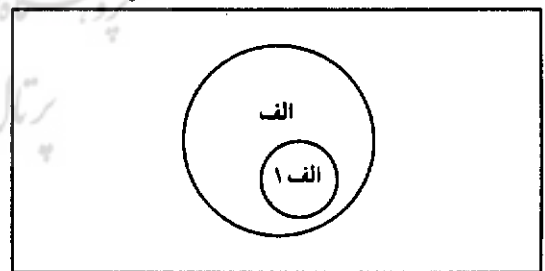
بازتابی نشده است.

این یک نتیجه نمونه است. در هر کاوش بسیاری از مدارک مجموعه بازتابی نمی‌شوند. به منظور یافتن نسبت مانعیت در کاوش باید در پایان چه تعداد از ۵۰ مدرک مرتبط است. این نسبت را می‌توانیم از طریق ارائه مدرک به شخص درخواست کننده و درخواست از او برای ارزیابی ارتباط مدرک با نیازهایش به دست آوریم. در یک مورد فرض درخواست کننده فقط ۱۰ مدرک را مربوط می‌داند بنابراین نسبت مانعیت $\frac{1}{50}$ یا ۲۰٪ است.

تعیین نسبت جامعیت کار ساده ای نیست. برای این منظور باید مشخص نماییم چه تعداد از مدارک بازتابی نشده در کاوش با درخواست کاوشگر مرتبط است. تنها یک راه برای تعیین دقیق جامعیت در یک کاوش وجود دارد و آن بررسی یکایک مدارک بازتابی نشده است. این امر مقدور نیست مگر در مورد وضعیت های آزمایش بسیار کوچک.

از آنجاییکه تعیین جامعیت مطلق عملاً میسر نیست تنها به برآوردی از جامعیت بسنده می‌شود. یکی از روشهای مطلوب در این مورد، مینا قرار دادن زیر مجموعه ای از مدارک برای برآورد جامعیت است که در پایگاه اطلاعاتی، مرتبط تشخیص داده شده است. (نمودار ۴)

مجموعه «ک»



نمودار ۴. روش برآورد جامعیت بر مبنای زیر مجموعه مدارک مرتبط موجود در پایگاه اطلاعاتی

برای هر تقاضا، مجموعه "ک" شامل یک مجموعه مدرک "الف" است که پرسشگر آنها را مشاهده می‌کند و بنابراین خود مرتبط می‌داند به عبارت دیگر این مجموعه "الف"، همان مجموعه $a+c$ است. (نگاه کنید به جدول ۱) بدین ترتیب می‌توانیم برآورد جامعیت را بر مبنای زیر مجموعه "الف ۱"



رعایت کل نگری را کرده ایم . بدین ترتیب هرگاه نمایه سازی کل نگر و در سطح عالی انجام گیرد ، جامعیت افزایش می یابد . از طرفی کل نگری و پیامد آن جامعیت ، موجب کاهش عملکرد مانعیت می گردد . این امر به دو دلیل اتفاق می افتد یکی به سبب اینکه اگر در هر مدرکی کلیه موضوعات را در نظر بگیریم در نتیجه حین نمایه سازی عنوانهای متعددی خواهیم داشت که تنها بخش اندک از مدرک بدان پرداخته و در نتیجه هنگام بازیابی آنها نیز علی رغم کمی اطلاعات ، مدارک بازیابی می شوند .

دوم آنکه با افزایش کلیدواژه های نمایه ، در مرحله کاوش ، همارایی کاذب (False Coordination) رخ می دهد و مآلاً روبه کاهش می گذارد (۳ ، ص ۶۶ - ۶۷) .

ب - جزء نگری (Specificity)

همان گونه که قابلیت های جامعیت یک نمایه از طریق سیاست گذاری در باب کل نگری نمایه سازی سنجیده می شود ، مانعیت یک نمایه کاملاً قائم به توانایی زبان نمایه سازی برای توصیف دقیق واژه هاست .

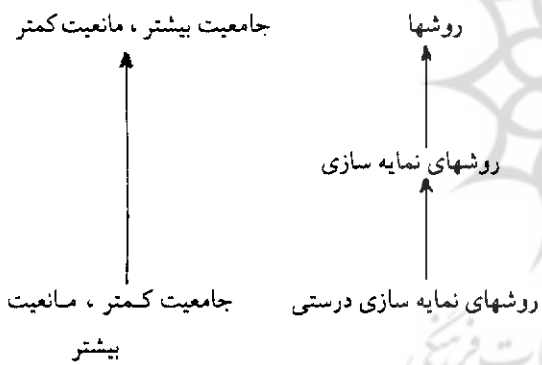
برای مثال می خواهیم مجموعه مدارکی را که در مورد بحث های مختلف نمایه سازی است نمایه کنیم . یکی از موضوعات اصلی در این مجموعه مدارک " روشهای نمایه سازی دستی " است که در نظام های فرضی الف ، ب ، ج مورد نمایه سازی قرار می گیرد . اگر در نظام الف که دارای ۲۰۰۰ واژگان نمایه ای است این مجموعه مدارک در مورد روشهای نمایه سازی دستی را نمایه سازی کنیم می توانیم دقیقاً واژه های روشهای ، نمایه سازی و دستی را وارد سازیم . اگر بخواهیم در نظام ب که دارای ۱۰۰۰ واژه است نمایه سازی کنیم ناچار باید فقط واژه های روشهای و نمایه سازی را به عنوان واژه های نمایه ای انتخاب نماییم . نهایتاً در نظام ج که دارای ۵۰۰ واژگان است فقط لفظ روشها را به عنوان خاص ترین روش می توان در نظر گرفت . نتیجه آنکه هنگامیکه تقاضایی برای جستجو در نظام الف در مورد روشهای نمایه سازی دستی ارائه می گردد می توانیم عنوان را به دقت همان گونه که در هنگام نمایه سازی وارد ساخته ایم ، کاوش نماییم .

و می توان انتظار داشت که مدارک بازیابی شده عمدتاً

مرتبط با موضوع است . از سوی دیگر نسبت جامعیت کاوش احتمالاً پایین خواهد آمد ، زیرا نظام ، مدارکی را که شامل اطلاعات مفیدی در مورد موضوع است نادیده خواهد گرفت . برای مثال با پرسش دقیق برای مجموعه مدارکی که با روشهای نمایه سازی دستی ارتباط دارد ما در بازیابی ، دو دسته مدارک را از دست خواهیم داد .

۱. مدارکی که با روشهای نمایه سازی به طور کلی مرتبط است و شامل اطلاعات اساسی و پایه در مورد روشهای نمایه سازی می گردد . این مدارک تحت واژه روشهای نمایه سازی نمایه می شوند .

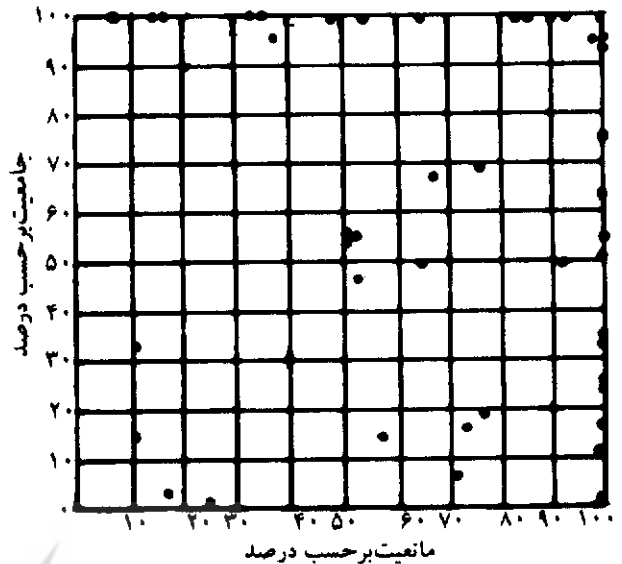
۲. مدارکی که با روشهای نمایه سازی دستی مرتبط است اما تحت واژه های مترادف یا ترکیب واژه ها می آید . نظیر شیوه های نمایه سازی دستی یا فنون نمایه سازی دستی (۳ ، ص ۶۸ - ۶۹)



جامعیت و مانعیت در کاوشهای نظام بازیابی چکیده های صرع : کاوش موردی

به منظور مشخص ساختن جامعیت و مانعیت در یک نظام بازیابی ، ۴۵ کاوش آزمایش در نظام بازیابی چکیده های صرع صورت گرفته که نمودار پراکندگی نتایج کاوشها در ذیل آمده است .

در نمودار ۵ . مجموعه ای از کاوشهای مطلوب وجود دارد که در قسمت بالای سمت راست نمودار قرار دارد (جامعیت کامل و مانعیت کامل) و همچنین چندین نمونه از کاوشهای نامطلوب (جامعیت کم ، مانعیت کم) در قسمت پایین سمت چپ نمودار ، و نیز نتایج کاوش جامعیت کامل و مانعیت کم و برخی کاوشهای با مانعیت کامل و جامعیت کم



نمودار ۵. پراکنندگی نتایج کاوش در نظام بازیابی چکیده های صرع

وجود دارد. کاوشهایی با جامعیت ۵۰٪ و مانعیت ۵۰٪ نیز در نمودار به چشم می خورد.

در جریان کاوشها و نیز مشخص شدن جامعیت و مانعیت چندین کاوش متفاوت، عوامل موثر در کاهش جامعیت و مانعیت چنین عنوان شده است:

کاهش در جامعیت

- الف. جامع نبودن روشهای گوناگون در بازیابی
- ب. بسیار خاص بودن استراتژی کاوش
- ج. بسیار عام بودن استراتژی کاوش

کاهش در مانعیت

- الف. نمایه سازی: کلی نگری در نمایه سازی
- ب. واژگان: همارایی کاذب، روابط نادرست اصطلاحات، ابهام متن
- ج. استراتژی کاوش: عدم جامعیت کاوش، عدم مانعیت کاوش، ساختار غیر منطقی بولی (۱، ص ۲۲۲ - ۲۲۹)

روشهای افزایش جامعیت و مانعیت در نظامهای بازیابی اطلاعات

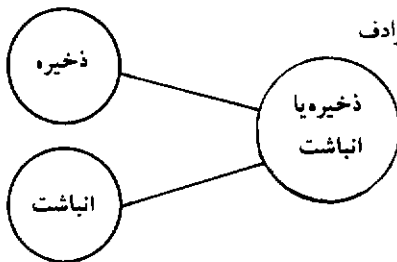
اصولاً در نمایه سازی تدابیری، برای به حد مطلوب

رساندن سطح جامعیت و مانعیت اندیشیده می شود.

الف. روشهای افزایش جامعیت. در واقع افزایش میزان رده های موضوعی مدرک، تعریف وسیع واژه ها و کاهش حجم واژگان از تدابیر اصلی در راستای افزایش جامعیت به شمار می آید و به طور اخص و به منظور ارائه روشهای علمی تر می توان از کنترل مترادفها و شبه مترادفها، کنترل شکل کلمات، پیوندهای سلسله مراتبی، دسته بندی و خوشه بندی (از حیث میزان همابندی دو یا چند واژه در یک مدرک) اشاره نمود.

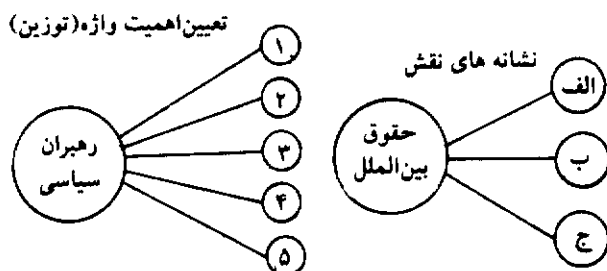
برای مثال:

نمونه کنترل مترادف



ب. روشهای افزایش سطح مانعیت. برای افزایش مانعیت نیز تدابیر گوناگونی وجود دارد. کاهش میزان رده های مدرک، محدود کردن تعریف واژه و افزایش حجم واژگان از بارزترین روشهای مطلوب سازی مانعیت به شمار می آید. مشخصاً همارایی، توزین (Weighting)، یا تعیین درجه اهمیت یک واژه در موقعیتهای مختلف، پیوند واژگان (Interlocking) که رابطه میان واژه های خاص را مشخص می سازد اما به کیفیت رابطه اشاره ای ندارد، شاخص های نقش (Role indicators) یا شاخص های ارتباطی که توسط نمایه ساز به اصطلاحات خاص جهت نشان دادن نقش دقیق یا کاربرد آن واژه در یک مدرک مشخص داده می شود، به طور کلی از ابزارهای افزایش میزان مانعیت به شمار می آید.

مثال





۲. عوامل مفید در مانعیت

ویژگی

در متن آزاد معمولاً واژگان خاص تر است در نتیجه مانعیت بالا می رود. لیکن تنوع روشهای ارائه مفاهیم، نیل به جامعیت بالا را در کاوشهای "مفهومی" عام مشکل می سازد. در کاوشهای از این نوع، واژه های کنترل شده نسبتاً عامتر، مرجح تر است.

عوامل مؤثر بر هر دو مقیاس

روز آمد بودن

عناصر برگرفته از متن آزاد همیشه روزآمدتر است. برای یافتن یک موضوع بسیار جدید در نظام واژگان کنترل شده ممکن است نیاز باشد کاوشگر واژه های متعددی را آزمایش کند (کاهش مانعیت) و نیز ممکن است در مورد آن موضوع چیزی نیابد (کاهش جامعیت)

شناخت

متخصصان اطلاع رسانی که با واژگان کنترل شده کاملاً آشنا هستند از این ابزار بهتر استفاده می کنند حال آنکه استفاده کننده نهایی با زبان طبیعی مدرک مورد نظرش آشنا تر است و بهتر کار می کند.

یکی از مطالعات انجام شده در زمینه مقایسه متن کامل، چکیده و نمایه سازی بر اساس واژگان کنترل شده توسط تنوپیر (Tenopir) در سال ۱۹۸۴ انجام گرفته است. با استفاده از بررسی بازگانی هاروارد بر روی خط Online، تنوپر به نتایج ذیل در مورد ۳۱ کاوش در این بانک رسیده است.

واژگان

کنترل

متن کامل چکیده شده

تعداد مدارک بازیابی شده (میانگین)	۱۷/۸	۲/۴	۳/۱
مدارک بازیابی شده مرتبط (میانگین)	۳/۵	۱/۰	۱/۲
جامعیت (در ارتباط با مجموع روشها)	۷۳/۹	۱۹/۳	۲۸/۰
مانعیت	۱۸/۰	۳۵/۶	۳۴/۰

ابزارهای افزایش مانعیت اساساً باعث افزایش حجم واژگان در نمایه می شود برای نمونه مثال بالا (توزین) به جای داشتن یک واژه پنج واژه داریم که با ارزشهای مفهومی متفاوت در موقعیتهای مختلف حضور دارند که این خود موجب افزایش حجم واژگان می گردد (۲، ص ۱۸۱).

تاثیر واژگان کنترل شده و زبان طبیعی بی جامعیت و مانعیت

الف. واژگان کنترل شده و تأثیر آن بر جامعیت و مانعیت

۱. عوامل مفید در جامعیت

طول مدرک (تعداد مجراهای بازیابی)

تاثیر نوع ارائه

اغلب عناصر متن آزاد (به جز عناوین)

طولانی تر از مجموعه ای از واژه های نمایه ای تعیین شده است. این امر موجب افزایش جامعیت و کاهش مانعیت می گردد.

حشرو زوائد

معمولاً در متن آزاد حشر بیشتر است و احتمال اینکه واژه ای خاص یافته شود نیز بیشتر است. با این حال روشهای گوناگونی برای نمایش یک موضوع در پایگاههای اطلاعاتی بزرگ متنی وجود دارد که نیل به جامعیت بیشتر را مشکل می سازد

وجود واژه هایی با "مفهوم" عام

بیشتر در واژگان کنترل شده ظاهر می گردد و ممکن است در متن به صورت ضمنی آمده باشد.

پیوند میان واژه هایی که از لحاظ معنا شناختی باهم مرتبطند

به وضوح در تشکیل واژگان کنترل شده با ساختار مناسب نقش دارد



برخی از افراد ممکن است ترکیبی از هر دو رانرجیح دهند. اما افراد معمولاً خواستار سطح "معقول" جامعیت و سطح "مقبول" مانعیت هستند.

یکی از محدودیتهای نسبت جامعیت این است که فرض را بر آن می گذارند که کلیه مدارک مرتبط با بازیابی شده تقریباً از ارزش یکسانی برخوردار است. این امر همیشه صادق نیست، در کاوشی که ۵ مدرک مرتبط را بازیابی کند و ۱۰ مدرک را از دست دهد (جامعیت ۳۳٪)، ممکن است ۵ مدرک بازیابی شده از حیث تناسب با نیاز بهتر از ۱۰ مدرک بازیابی نشده باشد. به طور کلی برای ارزیابی خدمات اطلاع رسانی، نسبت جامعیت باید به دقت و با ملاحظه به کار رود (۴، ص ۱۷۸).

یادداشتها:

* اصطلاحات "جامعیت" و "مانعیت" از برابر نهادهای آقای دکتر عباس خری برای دو اصطلاح Recall و Precision است.

- [1] Lancaster, F.W. Fayen, E.G. Information retrieval online. Los Angeles: Melville, 1973.
- [2] Lancaster, F.W. Indexing and abstracting in theory and Practice. London: the Library Association, 1991.
- [3] Lancaster, F.W. Information retrieval systems: Characteristics, Testing and evaluation. New York: John Wiley & sons, 1969.
- [4] Lancaster, F.W. "Precision and recall", Encyclopedia of library and information science, Vol. 23, PP. 170-188.
- [5] Pouitt, A.S. Information Storage and retrieval systems: Origins, development and application. New York: John Wiley & Sons, 1969.

شاید مهمترین نتیجه تحقیق تنویر را بتوان این دانست که کاوش با واژه های کنترل شده مدارکی را بازیابی می کند که در کاوش متن کامل بازیابی نمی شود و بالعکس.

در تحقیق دیگری مارتین بدین نتیجه رسیده است که عنصر زبان طبیعی پایگاه اطلاعاتی شامل تک واژه هایی است که توسط کامپیوتر از عناوین و چکیده ها استخراج شده و سپس توسط اشخاص مورد ویرایش قرار گرفته تا واژه های غیر اطلاعاتی حذف و واژگان را از طریق حذف املاهای متفاوت و اشکال مختلف کلمات مناسب سازد.

مارتین در جریان تحقیقات خود به نتایج ذیل رسیده است (۲، ص ۲۰۸، ۲۱۲):

جامعیت برحسب درصد مانعیت برحسب درصد

واژگان کنترل شده	۵۶	۷۴
زبان طبیعی	۷۸	۶۳
زبان طبیعی و واژگان کنترل شده	۷۱	۴۵

تنوع نیازهای استفاده کنندگان به جامعیت و مانعیت

افراد همیشه نیاز به جامعیت ندارند. استفاده کنندگان گوناگون، نیازهای متفاوتی به جامعیت و مانعیت دارند. حتی اهمیت جامعیت و مانعیت برای یک فرد خاص در زمانهای مختلف، متفاوت است. از یک سو تقاضاکننده ای داریم که به نگارش کتاب، تهیه مقاله یا انجام پروژه های تحقیقاتی بلند مدتی مشغول است. که احتمالاً خواستار جستجوی جامعی (جامعیت بیشتر) است و احتمالاً مانعیت، برای او کمتر اهمیت دارد. چون می خواهد مطمئن باشد که هیچ مدرک مهمی را از دست نمی دهد. از طرف دیگر استفاده کننده ای داریم که می خواهد از خدمات اطلاع رسانی صنعتی استفاده نماید و احتیاج به مقالات جدیدی در مورد موضوعی خاص دارد و سرعت تهیه آن برایش نیز حائز اهمیت است. این فرد احتیاج به جامعیت بالا ندارد اما در عوض مانعیت در نتایج کاوش بسیار او مهم است.