

اطلاع رسانی

فصلنامه علمی - پژوهشی مرکز اطلاعات و مدارک علمی ایران

دوره سیزدهم؛ شماره ۱. پاییز ۱۳۷۶

ریزش کاذب در ذخیره و بازیابی اطلاعات

نوشته: احمد یوسفی

کارشناس ارشد کتابداری و اطلاع رسانی

مرکز اطلاع رسانی و خدمات علمی جهادسازندگی

کلید واژه‌ها

ریزش کاذب، ذخیره و بازیابی اطلاعات، نمایه سازی، پایگاههای اطلاعاتی، نمایه سازی همارا، جامعیت، مانعیت

چکیده

مقاله حاضر تعریف، وجه تسمیه و علل ریزش کاذب، ریزش کاذب در نمایه‌سازی کوچک، کواک و پس همارا، شیوه‌های جلوگیری از ریزش کاذب، بویژه استفاده از "بط" و "نقش" و ارتباط ریزش کاذب با جامعیت و مانعیت را بیان می‌دارد. در پایان نیز ریزش کاذب در بانک‌های اطلاعاتی مورد بحث و بررسی قرار می‌گیرد.

تعاریف متعددی از ریزش کاذب^(۱) ارائه شده است که در اینجا به سه تعریف عمده و مهم اشاره می‌شود:

- "پایین افتادن کارتی که مرتبط و مورد نظر نیست در سیستم کارتی لبه منگنه"^(۲)
- "خروجی و برونداد غیرمرتبطی که در نتیجه اجرای راهبرد کاوش در سیستم بازیابی اطلاعات تولید می‌شود"^(۱)
- "اصطلاح فنی و تخصصی که برای اختلال در سیستم ارتباطی ذخیره و بازیابی اطلاعات به کار برده می‌شود."^(۱)

در نمونه اولیه چنین سیستم‌هایی، مدارک کدگذاری شده توسط کارهای هولریت^(۳)، که در آن کد مدارک بر روی کارت پانچ شده نشان داده می‌شد. سپس مرتب کننده کارت‌های پانچ شده به منظور جدا کردن کارت‌های مدارک مورد نظر، کارت‌های مرتبط را به داخل جای معینی پایین می‌انداخت. در میان مدارک خواسته شده، کارتی که نمایانگر مدرک ناخواسته بود و اشتباهاً

پایین افتاده بود، همان ریزش کاذب بود^{۱۳۱}. پس اصطلاح ریزش کاذب از اینجا ریشه می‌گیرد و گاهی به آن بازیابی ناخواسته^(۴) نیز می‌گویند.

علل ریزش کاذب

علت ریزش کاذب ممکن است افراد گوناگون باشند، یا در مراحل گوناگون اتفاق افتد، یا مربوط به خود سیستم باشد. نمایه‌ساز یا کدگذار ممکن است با تفسیر و تعبیر اشتباه از سیستم یا با اشتباه در ضبط و ذخیره‌نمایه یا کد، موجب ریزش کاذب شود؛ ممکن است هنگام اختصاص یک نمایه یا کد به مدرک، واژه یا اصطلاحی نامناسب برای مدرک انتخاب کند.

چنین اشتباهی در هنگام نمایه‌سازی یا کدگذاری نیز ممکن است اتفاق بیفتد.

در اینجا به کلیه عللی که ممکن است موجب اختلال و ریزش کاذب شود نگاهی می‌اندازیم:

۱- طراحی ضعیف و نامناسب نظام نمایه‌سازی و کدگذاری؛

۲- خطا در:

الف. نظام نمایه‌سازی و کدگذاری،

ب. انتقال نمایه و کد مدرک به رسانه ذخیره‌سازی،

پ. انتخاب نمایه یا کد در هنگام کاوش مسئله؛

۳- اختصاص نمایه و کد ضعیف و نامناسب به:

الف. مدرک، در رسانه ذخیره‌سازی،

ب. مسئله‌ای که باید کاوش شود؛

۴- تدوین راهبرد ضعیف و نامناسب در فرآیند کاوش؛

۵- اختلال در:

الف. عملکرد نظام ماشینی،

ب. خود مکانیسم؛

۶- ترکیب دو یا چند مورد از علل فوق‌الذکر؛^{۱۳۱}

۶- اشتباه کاربر و استفاده‌کننده در:

الف. انتخاب مجموعه یا پایگاه اطلاعاتی،

ب. جا انداختن مرحله‌ای از کاوش و بازیابی،

پ. هم‌آرایی توصیفگرها.

ریزش کاذب در نمایه‌سازی کوچک و کواک

چون این شیوه‌های نمایه‌سازی براساس کلمات موجود در عنوان مدارک صورت می‌پذیرند، دارای معایبی هستند؛ اما در اینجا فقط به آن معایبی که باعث ریزش کاذب می‌شوند اشاره می‌کنیم:

۱- ممکن است عنوان مدرک، بویژه در حوزه های علوم اجتماعی و انسانی، کاملاً گویای متن مدرک نباشد، یا به صورت استعاری بیان شده باشد. مثلاً اگر به دنبال مدرکی درباره «طلا» باشیم، چه بسا مدرکی که عبارت «طلای سیاه» در عنوان آن وجود دارد نیز بازایی شود که هیچ ربطی به «طلا» ندارد و درباره نفت است.

۲- ممکن است کلمات هم نویسه به اشتباه بازایی شوند؛ مثلاً کلمه «شیر» را کاوش کنیم و هدفمان مدارکی درباره شیر خوراکی باشد، ولی مدارک دیگری که در عنوان خود این کلمه را دارند ولی درباره «شیر آب» یا «شیر (حیوان)» هستند نیز بازایی شوند که طبعاً مورد نظر نبوده است.

۳- بازایی براساس تکواژه است و عبارت یا ترکیب را نمی توان کاوش کرد. (البته در سیستم کواک از عبارت نیز می توان استفاده نمود). مثلاً اگر درباره «کشاورزی ایران» مدرکی بخواهیم، چون بازایی به صورت تکواژه است باید هر یک از کلمات «کشاورزی» و «ایران» را جداگانه بکاوییم و کارت های آن ها را با یکدیگر همارایی کنیم؛ در این صورت نیز چه بسا مدرکی با عنوان «کشاورزی پاکستان و دامپروزی ایران» بازایی شود که مدرک دلخواه نیست و مصداق ریزش کاذب می باشد.^[۵]

ریزش کاذب در نمایه سازی پس همارا

در این شیوه همارایی هر یک از کلمات به طور مستقل بر روی برگه ای نوشته می شود و عمل همارایی را مراجعه کننده انجام می دهد. مراجعه کننده به راحتی می تواند هر کلمه را با کلمه یا کلمات دیگر همارایی کند و هیچ محدودیتی از این نظر ندارد. این اختیار، و نامحدود بودن همارایی کلمات، یکی از نقاط قوت این نوع نمایه سازی تلقی می شود. ولی به تجربه ثابت شده است که در مواردی ما را با ریزش کاذب مواجه می کند.^[۶]

مثلاً فرض کنید به دنبال مدرکی درباره «تأثیر محیط بر انسان» هستیم و کلیدواژه های ما عبارت اند از:

تأثیر/محیط/انسان. در هنگام بازایی ممکن است مدرکی به دست آید که هر سه کلیدواژه ها را دارد، ولی عنوان آن «تأثیر انسان بر محیط» باشد که با خواسته ما کاملاً متفاوت است. مثالی دیگر: به دنبال مدرکی درباره «فلسفه علم» هستیم، ولی در همارایی به مدرکی با عنوان «علم فلسفه» دست می یابیم که ارتباطی با خواسته و نیاز ما ندارد و ریزش کاذب است. برای جلوگیری از ریزش کاذب در سیستم پس همارا ۳ راه حل پیشنهاد می شود:

۱- پیش همارا کردن برخی از توصیفگرهایی که احتمال دارد ما را در هنگام بازایی دچار مشکل کند؛ در این راه حل چون عمل همارایی توسط نمایه ساز صورت می گیرد سعی می شود تا حد امکان از ریزش کاذب جلوگیری شود.

۲- استفاده از شاخص‌ها یا نشانه‌های «نقش»^(۵)؛ «نقش» برای نشان دادن عمل یا کارکرد یک توصیفگر معین در یک توصیف نمایه‌ای، و نشانه یا شاخصی است که کاربرد یک توصیفگر را در یک مدرک بخصوص مشخص می‌کند.^[7]

مثلاً ممکن است مدرکی با عنوان «انتظارات مردم از دولتمردان» داشته باشیم؛ توصیفگرهای این مدرک عبارت خواهند بود از: انتظارات/مردم/دولتمردان در هنگام بازیابی و هم‌ارایی ممکن است مدرکی با عنوان «انتظارات دولتمردان از مردم» نیز بازیابی شود که همهٔ توصیفگرهای آن مشترک اند، ولی مدرک مورد نظر ما نیست. برای جلوگیری از این ریزش کاذب می‌توان نقش توصیفگرها را در داخل پراتز در جلوی آن‌ها مشخص نمود و این‌گونه عمل کرد:

مردم (فعال)									
۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
				۵۴					

دولتمردان (غیرفعال)									
۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
				۵۴					

انتظارات									
۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
				۵۴					

۳- استفاده از «رابط»^(۶)؛ «رابط» نشانه یا نمادی است که برای پیوند دادن موضوع ها و مفاهیم توصیفگرهای یک مدرک بخصوص به کار می رود. رابط ها همراه با توصیفگرها و شماره مدارک مربوط به آن توصیفگرها می آیند تا ریزش کاذب را کم کنند.^(۸)

مثلاً اگر در مدرکی با عنوان «پرورش گوسفند و بیماری های بز»، توصیفگرها عبارت اند از: پرورش/گوسفند/بیماری ها/بز. در همارایی ممکن است توصیفگرهای «پرورش» و «بز»، یا توصیفگرهای «گوسفند» و «بیماری ها» با یکدیگر همارایی شوند، که در هر دو حالت با محتوای مدرک متفاوت است و ریزش کاذب خواهیم داشت. برای جلوگیری از این ریزش کاذب می توان بدین گونه عمل کرد:

رابط	شماره مدرک	توصیفگر
الف	۳۶	پرورش
الف	۳۶	گوسفند
ب	۳۶	بیماری ها
ب	۳۶	بز

این روش در کارت های ۹- بدین ترتیب صورت می پذیرد:

پرورش									
۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
						۳۶ الف			

گوسفند									
۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
						۳۶ الف			

بیماری ها									
۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
						۳۶ ب			

بزرگ									
۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
						۳۶ ب			

توضیحگر و ریزش کاذب

یکی دیگر از راه های جلوگیری از ریزش کاذب، استفاده از توضیحگر^(۷) است. زمانی که دو یا چند مقوله با کلمه ای ثابت بیان شوند، می توان از توضیحگر در داخل پرانتز استفاده کرد. مثلاً «اصفهان» هم نام استان است و هم نام شهر. برای تمایز این دو، برای یکی توضیحگر می آوریم:

اصفهان

اصفهان (استان)

مثال دیگر:

حرکت (بازیگری)

حرکت (فلسفه)

حرکت (فیزیک)

توضیحگر درون پرانتز و در مقابل موضوع یا توصیفگر، و معمولاً برای تعیین حدود معنی و برطرف کردن شبهه و اخص کردن موضوع می آید. توضیحگر در واقع به معنی «از نظر» است.

مثلاً حرکت از نظر فلسفه حرکت (فلسفه)؛

یا حرکت از نظر فیزیک حرکت (فیزیک)^{۹۱}

ارتباط ریزش کاذب با جامعیت و مانعیت

معمولاً ریزش کاذب درصد کمی از مدارک بازیابی شده را شامل می شود؛ چرا که برخی از خطاهای بازیابی گریز ناپذیرند، بویژه هنگامی که به منظور دستیابی به جامعیت بالا، کاوش را گسترده کنیم. یعنی هر چه کاوش را گسترده تر کنیم و بخواهیم جامعیت را بالا ببریم، در مانعیت دچار مشکل می شویم و رکوردهای ناخواسته بازیابی شده زیاد می شود که همان ریزش کاذب است. در نتیجه، هر چه میزان جامعیت بیش تر باشد، میزان ریزش کاذب نیز زیاد می شود و هر چه میزان مانعیت زیاد باشد، میزان ریزش کاذب کاهش می یابد.

بنابراین، فرمول های زیر برای ریزش کاذب، جامعیت و مانعیت ارائه می شود:

$$\text{ریزش کاذب} = \left\{ \text{مدارک بازیابی شده مرتبط} \right\} - \left\{ \text{تعداد کل مدارک بازیابی شده} \right\}$$

ریزش کاذب - تعداد کل مدارک بازیابی شده = مانعیت

$$\text{جامعیت} = \left\{ \text{تعداد توصیفگرهایی مورد نظر} \right\} - \left\{ \text{تعداد کل توصیفگر مورد بازیابی در پایگاه اطلاعاتی} \right\}$$

ریزش کاذب در بانک های اطلاعاتی

ریزش کاذب در بانک های اطلاعاتی ممکن است به علل زیر اتفاق بیفتد:

- ۱- اشتباه و خطا در بیان پرسش، که مهم ترین مرحله است؛ زیرا اگر در این مرحله اشتباه صورت پذیرد، در بقیه مراحل نیز دچار خطا خواهیم شد.
- ۲- اشتباه در تحلیل پرسش؛
- ۳- اشتباه در انتخاب فایل و پایگاه اطلاعاتی مناسب؛

- ۴- اشتباه در تدوین راهبرد کاوش؛
- ۵- اشتباه در ساختار زبانی و ساختار کلمه^{۱۰۱}؛
- ۶- استفاده از زبان کنترل شده (در موارد استثنایی که متعاقباً ذکر می شود)؛
- ۷- استفاده از کلیدواژه کنترل نشده؛
- ۸- کاوش کلمه ای در کلیه فیلدها، فیلد آزاد، یا متن آزاد^(۸)؛
- ۹- نبودن اصطلاحنامه در پایگاه اطلاعاتی یا ضعیف بودن آن؛
- ۱۰- بازیابی ناخواسته هم نویسه ها، بویژه در سرنام ها؛ مانند: Aids و AIDS که اولی به معنی وسایل و ابزار کمکی است و دومی حروف آغازین کلمات Acquired Deficiency Syndrome Immune یا SDI که هم مختصر Selective Dissemination of Information است و هم مختصر Strategic Defense Initiative.
- برای جلوگیری از ریزش کاذب در بانک های اطلاعاتی، استفاده از شیوه های زیر توصیه می شود:
- ۱- آشنایی کامل با بانک اطلاعاتی مورد نظر، شیوه های کاوش، پوشش موضوعی و امثال آن؛
- ۲- استفاده از عملگر AND برای اخص کردن مدارک بازیابی شده و افزایش مانعیت؛
- ۳- استفاده از عملگر NOT برای حذف رکوردهای ناخواسته؛
- ۴- استفاده از زبان کنترل شده و اصطلاحنامه؛
- ۵- کاوش کلمه مورد نظر در متن آزاد و بدون استفاده از زبان کنترل شده (در موارد استثنایی). معمولاً استفاده از زبان کنترل شده باعث رفع بسیاری از ریزش های کاذب می شود. با وجود این، اطلاعات جدید و بسیار جزئی و اخص، که معمولاً اصطلاحنامه و کنترل واژگان وجود ندارد، اغلب اوقات فقط از طریق کاوش کلمه در متن قابل بازیابی و دستیابی هستند و اگر در چنین مواردی از زبان کنترل شده استفاده کنیم، ریزش کاذب خواهیم داشت.^[۱۱]

پی نوشت ها

1. False drop
2. Edge-notched card system
3. Hollerith cards
4. Unwanted retrieval
5. Role
6. Link
7. Qualifier
8. Free text

منابع

- [1]. Bose, Anindya (1986). "Information resources management: a glossary of terms" in: Encyclopedia of Library and Information Science, v.41, p.116.
- [2]. Newman, simon (1997) M. "Noise" In : Encyclopedia of Library and Infomation Science, V.20, p.69.

[3]. Ibid., p.69.

[4]. Ibida., p.69-70.

[۵]. صدیق بهزادی، ماندانا. جزوه درسی ذخیره و بازیابی اطلاعات. نیمسال دوم ۷۶-۷۵، دانشکده روانشناسی و علوم تربیتی دانشگاه تهران.

[۶]. آقا بخشی، علی (۱۳۷۲). نمایه‌سازی همارا. تهران: مرکز اطلاعات و مدارک علمی ایران؛ ص

۱۲.

[7]. Rowley, Jennifer E. (1982). Abstracting and Indexing. 2nd ed, London: Clive Bingley, p.124.

[8]. Brown, A.G. (1982). An introduction to subject indexing. 2nd ed, London: Clive Bingley, p.378.

[۹]. سلطانی، یوری و کامران فانی ۱۳۷۳. سرعنوان‌های موضوعی فارسی. ویرایش ۲، تهران:

کتابخانه ملی جمهوری اسلامی ایران ص. بیست و هشت.

[10]. Shuman, B.A. (1992). "One false drop deserves another: file selection as a means of increasing precision in online searches". In: Proceedings of the 13th National Online Meeting, New York, May 1992, New Jersey: Learned Information, pp.345-350.

[11]. Wagner, E. (1986). "False drops: how they arise... how to avoid them" "online retrieval". Online, v.10, NO.5, pp.93-97.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی