

# جایگاه نام نویسندگان در بازیابی اطلاعات پایگاه‌ها



مصطفی علیمرادی \*

malimoradi@noornet.net

## چکیده

نویسنده اثر، از دیدگاه‌های گوناگون بر اثر تأثیر دارد و گاه دقیق نبودن یا مبهم بودن نویسنده یک اثر، سبب مهجور شدن آن در میان متخصصان علم می‌شود. از همین روی، در ادوار گوناگون، فهرست کردن آثار نویسندگان و ثبت و ضبط آنها، مورد اهتمام بزرگان علم بوده است. در این نوشته، با توجه به اهمیت نویسنده اثر در شناخت درست محتوای آن، به بررسی برخی فعالیت‌های صورت گرفته در پایگاه‌های مرکز نور برای نظام‌بخشی به نویسندگان اثر در بازیابی محتوا و نیز ناهمسانی‌ها و کاستی‌های موجود در ثبت نام نویسندگان در این پایگاه‌ها می‌پردازیم و راه‌هایی برای حل این معضل پیشنهاد می‌دهیم.

**کلیدواژگان:** نویسنده، فهم اثر، نام مستعار، تشخیص موجودیت اعلام، محصولات مرکز نور، بازیابی اطلاعات.

\* پژوهشگر مرکز تحقیقات کامپیوتری علوم اسلامی.

## اهمیت نام نویسنده اثر در جامعه علمی

از عناصر بسیار مهم در شناخت یک منبع، نام نویسنده آن است. شناخت نام نویسنده اثر، در تأویل و تفسیر آن اثر نیز بسیار مهم شمرده می‌شود. یک خطا در اسناد اثر به نویسنده‌ای خاص، شاید تفسیر آثار دیگر او را نیز تحت الشعاع قرار دهد و سبب چشم‌پوشی از اصل محتوا و مقصود نویسنده در آثار دیگرش شود؛ برای نمونه، بسیاری از حکمای اسلامی به اشتباه کتاب «اثولوجیا» را که اثر «فلوطین» است، به «ارسطو» نسبت می‌دادند؛ این در حالی است که فلوطین از پیروان افلاطون شمرده می‌شد و گرایش‌های شهودی که رنگ‌وبوی عرفانی داشت، در آثارش غالب بود؛ اما ارسطو بیشتر متکی به عقل‌گرایی اعتدالی بود؛ ولی از آنجا که دیدگاه فلوطین در اثولوجیا، بیشتر به مذاق حکمای مسلمان خوش می‌آمد، گاه ارسطو را حکیمی دارای مذاق الهی می‌خواندند.

«و همچنین ارسطاطالیس معلم اول و سرسلسله حکمای مشائین نیز این صور قدسیه الهی را مشاهده کرده است؛ چنان که کتاب او که به نام اثولوجیا معروف است، بر این مشاهده دلالت دارد [و مشاهدات خویش را در آن کتاب بیان می‌کند].» (سجادی، ۱۳۷۹، ۴۹۱)

و برخی دیگر گفته‌اند:

«و در طی این، سرّ دقیقه به‌غایت دقیق و لطیف دانستنی است که معلم مشائیه یونانیین، ارسطاطالیس، در کتاب اثولوجیا بر آن تنبیه کرده و ما در کتب و صحف حکمیه خود، طسقی صالح، و قسطی وافر از حق کلام در شرح و بسط آن آورده‌ایم.» (داماد، ۱۳۸۰، ۱۵)

شاید اگر حکمای مسلمان می‌دانستند، اثولوجیا از آن فلوطین است، می‌کوشیدند آثار دیگر او را نیز بیابند و مسیر تفکر، به گونه‌ای دیگر رقم می‌خورد.

همچنین، در تاریخ عرفان و ادب فارسی، نبود دقت در ثبت و خوانش کامل نام‌ها، سبب شده است بسیاری از منابع میان کسانی که نام مشترک یا نزدیک به هم دارند، خلط شود؛ از آن نمونه، می‌توان به خلط آثار «ابوحامد غزالی» که متکلم است با «ابوحامد غزالی» که فقیه و صوفی شمرده می‌شود، اشاره کرد که در منابع و نوشته‌های بسیار، میان آنها خلط شده است. چنانکه اگر بخواهیم نام کامل آنها را بنویسیم، باید بگوییم: «ابوحامد محمد غزالی»، متکلم است و آن که فقیه و صوفی است، «ابوحامد احمد غزالی» است. بر این اساس، شاید بتوان گفت اهتمام به ثبت دقیق و کامل نام پدیدآورندگان آثار، نقشی بسیار پررنگ و اثرگذار در تاریخ اندیشه و فرهنگ ایفا می‌کند.

افزون بر نمونه‌های یادشده، امروزه وجهی دیگر برای ثبت دقیق و کامل نام مؤلفان نیز مطرح می‌شود و آن، حقوق معنوی اثر است. اثری که تولید می‌شود، ممکن است آثار گوناگون مثبت و منفی بر جامعه بگذارد و نویسنده باید پاسخگوی این پیامدها باشد. همچنین، در صورت سوء استفاده از اثرش، بتواند به شکل قانونی پی‌گیر باشد. بنابراین، ثبت نام واقعی و کامل مؤلف، از این جهت نیز بایسته می‌نماید.

پس از بیان مقدمات، ضرورت طرح مسئله و جایگاه نام نویسندگان در بازیابی اطلاعات عرضه می‌گردد و پس از آن، فعالیت‌های مرکز در نظام‌مند کردن نام نویسندگان در پایگاه‌های کتابخانه‌ای و نواقصی که در آن ممکن است وجود داشته باشد، بررسی می‌شود و راه‌حلهایی برای هریک از این مسائل بیان خواهد شد. در پایان نیز پیشنهادهایی برای سامان‌بخشی به نام نویسندگان آثار عرضه خواهد گردید.

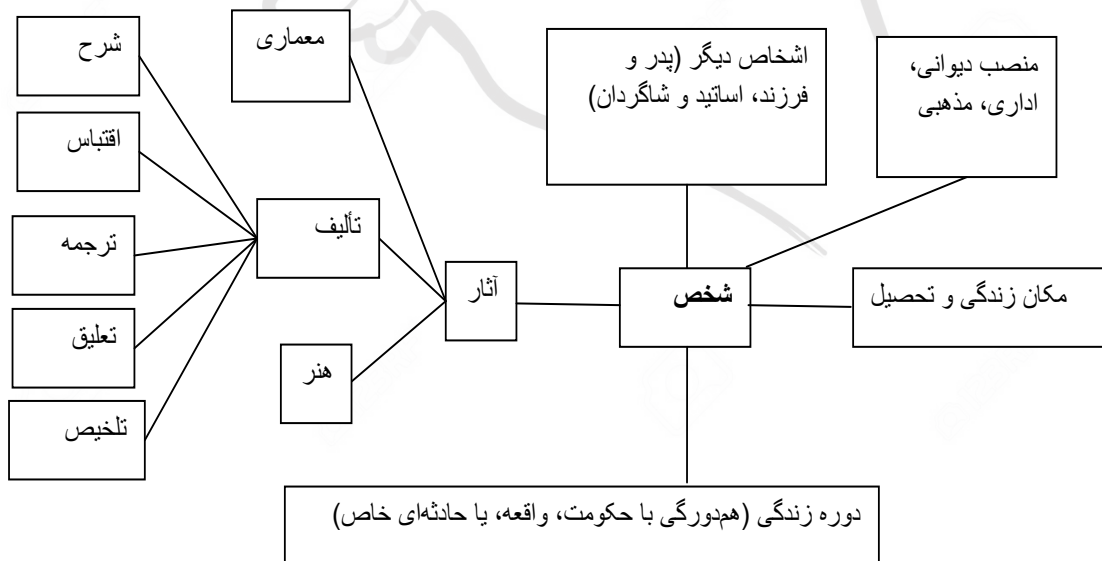
## ضرورت حل مسئله نام نویسندگان

۱. نام نویسنده مقاله، از راه‌های رسیدن به محتواست: نام پدیدآورندگان، از راه‌های دستیابی به محتوای پایگاه است؛ به این معنا که کاربران گاه تنها اطلاعاتی که از یک مقاله دارند، نام نویسنده آن است و جز آن، هیچ کلیدواژه‌ای که آنها را به مقاله مزبور رهنمون سازد، در دست ندارند. اگر نام نویسنده به‌درستی و کامل ثبت نشده باشد، این کاربر به مقاله نخواهد رسید.

۲. ثبت صحیح و کامل نام نویسنده مقاله، زمینه‌ساز ایجاد نظام تشخیص موجودیت نام (Named Entity Recognition (NER)) برای تسریع و ایجاد زمینه‌هایی برای رسیدن به هستی‌شناسی علوم اسلامی است.

امروزه، از حوزه‌های بسیار مهم در استخراج اطلاعات (information extraction)، تشخیص وجودیت اعلام است که با استفاده از فنون پردازش زبان طبیعی (natural language processing) انجام می‌شود. تشخیص موجودیت اعلام در پی یافتن و طبقه‌بندی عناصر اعلام در یک متن، بر پایه طبقه‌بندی‌های از پیش تعیین شده مانند: نام اشخاص، سازمان‌ها، مکان‌ها و زمان است. (Nadeau and Sekine. 2007, 18)

در نظام‌های بازبایی اطلاعات، نام‌ها افزون بر آنکه خود واحد اطلاعاتی بسیار مهم در یک متن به شمار می‌روند، روشن‌کننده اطلاعات بسیار دیگر نیز هستند که برای کاربران خیلی مهم است؛ زیرا نام‌ها به همراه خود، عناصر اطلاعاتی بسیار دیگر به جویندگان عرضه می‌کنند. یک نام اگر به‌درستی تبیین شود، از زمان، مکان، آثار و اشخاص دیگر نیز آگاهی‌های بسیار مفید به جوینده اطلاعات عرضه می‌کند. هنگامی که نام به گونه کامل و دقیق ثبت شود، دست‌کم نخستین چیزی که همراه آن روشن می‌شود، آثارش است؛ اما در نمونه‌های بسیار، مواردی همچون: تاریخ تولد، مکان زندگی، کار، تحصیل، افراد همکار، شاگردان و اساتید، را نیز روشن می‌کند. بنابراین، نام نویسنده از آنجا که به انسان مرتبط است و در دنیا منشأ اثر واقع می‌شود و فعالیت‌هایش بر دیگران تأثیر می‌گذارد و از ایشان تأثیر می‌پذیرد، خود سبب ایجاد وقایعی می‌شود که آگاهی از آن برای جوینده اطلاعات بسیار مفید است. از سوی دیگر، چون هر شخص، یکه و یگانه است و هر شخصی دنیای خود را دارد و هیچ‌گاه نمی‌توان آن را در یک امر منجذب ساخت و از شناخت هر شخص یگانه بی‌نیاز شد، تمیز دادن نام‌ها و یافتن صاحب نام واقعی سبب دقت اطلاعات خواهد شد. همچنین، به سبب همین یگانگی اشخاص، شکل‌های متکثر و متعدد ضبط نام‌ها نیز باید به همین اشخاص تقلیل یابد تا ارزش اطلاعاتی را در بازبایی محتوا داشته باشد.



شکل ۱: شبکه اطلاعاتی که ممکن است حول یک شخص گرد آید.

در این نمودار، ارتباط‌هایی که شخص به منزله «نویسنده» دارد، تولید یک اثر است و اثر، یا تألیف است که شاید ترجمه، تعلیق، تلخیص و یا شرح داشته باشد و به این واسطه با آنها نیز مرتبط می‌شود، یا اثر نویسنده، ترجمه، شرح یا تعلیقه برای یک اثر دیگر است که با آن اثر مرتبط خواهد شد.

همچنین، نویسنده به منزله یک انسان، والدین و فرزندان، مکان زندگی و تحصیل، و دوره زمانی زندگی نیز دارد و آگاهی‌هایی از آنها نیز می‌دهد. از این‌رو، نام‌های اشخاص، از مهم‌ترین واحدهای اطلاعاتی به شمار می‌روند.

### بررسی نام پدیدآورندگان در پایگاه‌های مرکز

مرکز تحقیقات کامپیوتری علوم اسلامی، بیش از ده پایگاه فعال دارد که در هر یک هدفی خاص پی‌گیری می‌شود و در میان این پایگاه‌ها، دست‌کم پنج پایگاه کتابخانه‌ای است و به عرضه محتوای پژوهشی می‌پردازد؛ یعنی پایگاه‌های: کتابخانه دیجیتال نور، مجلات تخصصی نور، حوزه، فصلنامه ره‌آورد نور و پایگاه نورحدیث. در این مقاله، تنها به بررسی امکانات و قابلیت‌های به‌کاررفته در نظم‌بخشی به نام پدیدآورندگان و برشمردن مشکلات آن در سه پایگاه اصلی: حوزه، کتابخانه دیجیتال نور و مجلات تخصصی پرداخته خواهد شد.

#### الف - نام پدیدآورندگان در پایگاه حوزه

در پایگاه حوزه ([www.hawzah.net](http://www.hawzah.net)) چندگونه محتوا عرضه می‌شود که شاید بتوان به گونه کلی، آنها را در چهار دسته جای داد:

کتاب، بخش‌هایی از کتاب، مقالات (شامل: مقالات مجلات، مقالات منتشرشده در پایگاه‌های دیگر که در پایگاه حوزه بازنشر شده، مقالات تولیدی محققان برای پایگاه) و چکیده پایان‌نامه‌ها.

از میان این محتواها، تنها نام نویسندگان کتاب‌ها و پایان‌نامه‌ها در پایگاه ذکر شده است.

برای مقالات، به گونه معمول نام پدیدآور نیامده است؛ در صورتی که نام نویسنده مقالات، اعم از شخص حقیقی و یا نام پایگاه، باید در پایگاه آورده شود تا هم استناد به آن آسان باشد و هم حق مالکیت معنوی نویسندگان حفظ شود.

#### ب - نام پدیدآورندگان در پایگاه کتابخانه دیجیتال (نور لایب)

در پایگاه کتابخانه دیجیتال نور ([www.noorlib.ir](http://www.noorlib.ir))، زبانه‌ای مجزا برای پدیدآورندگان تعبیه شده است که با انتخاب آن، کاربران می‌توانند از دو طریق مرور الفبایی و جست‌وجو، به نویسندگان دلخواه دسترسی داشته باشند. همچنین در این پایگاه، نقش کسان در پدیدآمدن اثر نیز مشخص شده است؛ برای نمونه، پس از مرور یا جست‌وجوی نام پدیدآوران و در صفحه نتیجه، هنگامی که روی نام پدیدآورنده کلیک شود، عنوان کتاب‌های مرتبط با او نمایش داده می‌شود و نقش او در پدیدآمدن اثر، از قبیل: نویسنده، مترجم، محقق، حاشیه‌نویس، مقدمه و تصحیح نیز مشخص می‌شود.

اگرچه کوشش شده در این پایگاه شیوه ضبط نام‌ها به شکل یک‌دست و یکسان باشد، اما هنوز مشکلات بسیاری در نام پدیدآورندگان در این پایگاه دیده می‌شود؛ از آن جمله می‌توان به نمونه‌های ذیل اشاره کرد:

۱. یکسان نبودن نام نویسندگان از لحاظ ثبت کردن یا ثبت نکردن پسوند نام‌ها؛

برای نمونه، یک نام گاه به شکل «ابن بابویه، علی بن حسین» آمده و در جایی دیگر به صورت «ابن بابویه قمی، علی بن حسین» ثبت شده است.

۲. رعایت نشدن یک‌دستی در ثبت نام مؤلفان هنگام ورود اطلاعات به لحاظ افزودن یا کاستن حرف یا نویسه‌ای خاص، گاه به سبب سهو اپراتورها یا تغییر زبان صفحه‌کلید، با وجود یکسانی ظاهری در نام پدیدآورندگان، ماشین، یک پدیدآورنده خاص را چند پدیدآورنده بر شمرده است.

در خصوص راه حل مسئله نام نویسندگان در پایگاه کتابخانه دیجیتال باید گفت، شاید تعیین ملاک برای ضبط نام، بر عهده نهادی خاص است و همه نهادهای علمی باید از یک رویه واحد برای ثبت نام نویسندگان استفاده کنند. مراکزی چون سازمان کتابخانه ملی ایران، از این دست نهادها شمرده می‌شوند. مستند مشاهیر، فرهنگ اعلام فارسی و کتاب‌شناسی ملی ایران، از جمله مستندات اند که می‌توان با استفاده از آنها، ملاکی یگانه در ثبت نام نویسندگان در دست داشت.

بر پایه اشکال‌های یادشده، شاید حل معضلات و اتخاذ اسلوب و الگویی واحد در ثبت نام نویسندگان، ضروری باشد. برای این کار، شایسته است که مسئولان ورود اطلاعات منابع، همه مدیران را ملزم به رعایت استاندارد پذیرفته در ثبت نام‌ها کنند. این استانداردها، چنانکه گذشت، از طریق کتاب‌شناسی ملی ایران یا اسنادی که سازمان‌های مسئول در دست دارند، قابل استخراج است.

از آنجا که صورت‌های گوناگون ضبط نام نویسندگان در برخی نرم‌افزارها به هم ارجاع شده‌اند، با اصلاح یک نام، صورت‌های دیگر آن نیز به گونه خودکار امکان اصلاح دارد.

برای آنکه شیوه ثبت نام نویسندگان در پایگاه منطبق بر معیارهای پذیرفته باشد، این کار باید بر پایه کتب مرجع پیش گفته صورت گیرد.

برای حل معضل ناهمگونی نام‌ها به سبب فزونی و کاستی نویسه‌ها نیز می‌توان با یک دستور کلی به ماشین، این اشکال‌ها را حل کرد.

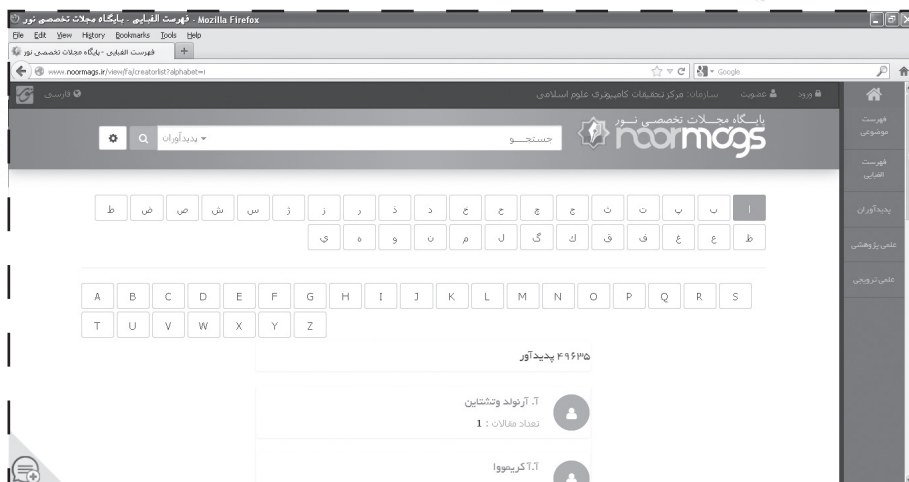
### ج - نام پدیدآورندگان در پایگاه مجلات تخصصی نور

از مهم‌ترین قابلیت‌های عرضه‌شده در پایگاه مجلات تخصصی نور ([www.noormags.ir](http://www.noormags.ir))، امکان جست‌وجو در نام پدیدآورندگان مقالات است. با استفاده از این امکان، کاربران به‌آسانی می‌توانند نویسنده دلخواه خود را در نوار جست‌وجو وارد کنند و همه مقالات او را در صفحه نتایج مشاهده کنند.

همچنین، امکان مشاهده نویسندگان با مقالات مرتبط نیز در پایگاه گنجانده شده است که بر پایه آن، نویسندگانی که در موضوع‌های مشابه مقاله نوشته‌اند، در ستونی مجزا به کاربران نمایش داده می‌شوند.

افزون بر آن، هنگام جست‌وجو در پدیدآورندگان مقالات، نویسندگان همکار نیز در ستونی مجزا به کاربر عرضه می‌شوند. مقصود از نویسندگان همکار، مترجمان، مؤلفان همکار یا نویسنده یک اثر است. از این رو، در مقالاتی که ترجمه‌اند، پدیدآوری که در صفحه نتایج آمده، مترجم مقاله، یا نویسنده همکار، مولف خواهد بود.

این اطلاعات عرضه‌شده برای کاربر، بسیار سودمند است و کار بازیابی اطلاعات را برای ایشان، دقیق‌تر و آسان‌تر خواهد کرد.



## کاستی‌های نام پدیدآورندگان در پایگاه مجلات تخصصی نور

نام پدیدآورندگان مقالات در پایگاه مجلات تخصصی نور (نورمگز)، مشکلات بسیار برای کاربران به وجود آورده است که از آن میان، می‌توان به مشتبه‌شدن مقالات نویسندگان هم‌نام، آمدن نام نویسنده واحد به اشکال گوناگون، و خطا در ثبت نام خانوادگی و نام در برخی نویسندگان اشاره کرد.

برخی از این معضلات، هنگام ورود اطلاعات و به واسطه خطای اپراتورهای خود پایگاه به وجود می‌آید که این لغزش‌ها را می‌توان به شکل نظام‌وار (سیستمی) برطرف کرد؛ اما عمده خطاها، از اطلاعات عرضه‌شده در مقالات مجلات ناشی می‌شود. حل این معضل، به مراتب مشکل‌تر است و به برنامه‌ریزی طولانی‌تر نیاز خواهد داشت.

بنابراین، منشأ خطا در نام‌ها و به‌تبع آن، بروز مسئله ابهام و اشتباه در نویسندگان مقالات را می‌توان در دو دسته کلی جای داد:

### ۱. خطا در ورود اطلاعات:

برخی لغزش‌های موجود در نام‌ها، به سبب خطایی است که از مجریان ورود اطلاعات سر می‌زند؛ به گونه‌ای که هنگام ثبت نام مؤلف، حرفی از آن می‌کاهند یا به آن می‌افزایند و یا نویسه (کارکتری) را به آن می‌افزایند یا از آن می‌کاهند. به این ترتیب، یک نویسنده با کاستن و افزودن یک حرف یا نویسه، به منزله دو نویسنده به شمار می‌رود.

در برخی از نمونه‌ها، نام و نام خانوادگی مؤلف به گونه پس و پیش نوشته می‌شود؛ برای مثال، نام یکی از پژوهشگران منطق، «عسگری سلیمان امیری» است که گاه به خطا «سلیمان امیری عسگری» نوشته می‌شود. گاهی نیز تعدد نام واحد، سببی دیگر دارد؛ به این معنا که گاه ظاهر نام نویسنده واحد شبیه هم است؛ اما باز ماشین آن را دو نویسنده مجزا تشخیص داده است. این به آن سبب است که زبان یا قلم نگارش آن دو، متفاوت بوده است. توضیح آنکه گاه ماشین به زبان نگارش برخی حروف (مانند «ی» و «ک») حساس است؛ اگر یک اپراتور با صفحه‌کلید فارسی یک نام را ثبت کند و اپراتور دیگر با صفحه‌کلید عربی همان نام را ثبت کند، ماشین این نویسنده واحد را به‌منزله دو نویسنده تشخیص می‌دهد و مقالات مجزا را ذیل نام آنها مندرج می‌کند.

برای حل معضلاتی که از این طریق به وجود آمده، چند راه حل ساده وجود دارد:

دقت در ثبت اطلاعات مجلات؛ بازنگری اطلاعات ثبت‌شده؛ دستور کلی به ماشین برای چشم‌پوشی از فاصله و نیم‌فاصله‌ها، دستور کلی به ماشین برای چشم‌پوشی از عربی و فارسی بودن نویسه‌ها.

### ۲. انتقال خطاهای موجود در مجلات به پایگاه:

گاه در مجلات، نام نویسنده مقالات به گونه ناقص یا نادرست ثبت می‌شود. این مسئله در مجلاتی که دارای رتبه علمی نیستند یا در دوره‌ای خاص منتشر شده‌اند، بیشتر به چشم می‌خورد.

به گونه کلی، معضل نام‌ها در پایگاه که به سبب نقص ثبت نام مؤلفان در مجلات پیش آمده، شامل نمونه‌های ذیل است:

– مجلاتی در آنها مقالاتی از نوع مصاحبه وجود دارد و نویسنده مقاله، مصاحبه‌شونده است و از این رو، نام مصاحبه‌شونده به سبب شهرت به گونه ناقص ثبت شده؛ مانند ذکر نام «استاد دینانی» به جای «غلام‌حسین ابراهیمی دینانی». یا به دلایل دیگر، فقط نام خانوادگی یا تنها نام مصاحبه‌شونده آورده می‌شود.

– مجلاتی که در اوضاع اجتماعی و سیاسی خاص منتشر می‌شده و نویسنده نام مستعار ذکر می‌کرده یا نام کوچک خود را به‌اختصار نام خانوادگی خود را به گونه کامل ثبت می‌کرده است؛ مانند نوشتن «د. ربیحی شرابی» به عنوان نام نویسنده.

- مجلاتی که در دوره گذشته منتشر می‌شده است و نوشتن نام و نام خانوادگی، به گونه اختصاری مرسوم بوده است؛ برای نمونه، «ع.ص» به جای «علی صفایی».

- مجلاتی که در آن قطعات ادبی یا اشعاری آمده و نام مؤلف با تلخیص شاعرانه ذکر گردیده است؛ مانند «ه. ا. سایه» به جای «هوشنگ ابتهاج».

- مجلاتی که رتبه علمی ندارند و محتوا، سبک نگارش، معیار و استاندارد برای نگارش مقالات در آن وجود ندارد و نویسندگان مقالات هیچ الزامی برای رعایت ملاکی خاص برای نشر مقالات در آن ندارند. در این مجلات، مقالاتی بدون ثبت نام نویسنده، نام تنها، یا نام خانوادگی تنها، نام مستعار و نام اختصاری، به چشم می‌خورد.

### \* راهکارهای حل مسئله نام در مجلات بدون رتبه علمی

از راه‌های پیش‌گیری از به وجود آمدن مسئله نام پدیدآورندگان، در مرحله ثبت و ورود اطلاعات انجام می‌شود؛ به این معنا که در ثبت نام پدیدآورندگان، از متخصصان کتاب‌شناس و مؤلف‌شناس سود برد یا در نمونه‌های پیش‌گفته با ایشان مشورت کرد. این مشورت ممکن است با شخص حقیقی خاص یا مؤسسه یا نهادی ویژه صورت گیرد و نیز می‌شود از طریق جست‌وجو در پایگاه‌های معتبر، مانند کتابخانه ملی، ثبت نام مؤلفان به‌درستی صورت گیرد. گاه گرفتن اطلاعات از نشریه‌ای که این مقالات در آن منتشر شده نیز کارساز خواهد بود.

افزون بر آن، برخی نام‌ها را می‌توان با استفاده از مستند مشاهیر یا فرهنگ اعلام فارسی اصلاح کرد. اما حل مسئله درباره مقالاتی که اطلاعات آن پیش‌تر ثبت شده، بسی دشوارتر خواهد بود؛ زیرا نخست باید معلوم کرد که چه تعداد از مقالات چنین معضلی دارند و این خود، بسیار زمان‌بر و دشوار است. در نمونه‌هایی که نویسنده از نام مستعار استفاده کرده، این یابش به‌سادگی قابل انجام نیست؛ مگر اینکه نام مستعار بسیار مشهور و شناخته شده باشد؛ مانند «نیما یوشیج» که نام مستعاری برای «علی اسفندیاری» است. شاید مناسب‌ترین کار آن باشد که تنها مجلاتی که محتوای غنی و پر بار دارند، مانند برخی نشریات قدیمی که افراد شاخص در آن مقاله منتشر کرده‌اند، برگزیده شود و نام پدیدآورندگان بررسی گردد و مشکلات، احتمالی نام‌ها به شیوه‌هایی که پیش‌تر گفته شد، رفع گردد. با این کار، نخست دایره مشکلات محدود به مقالات خاص می‌شود و نیز یافتن نام کامل و دقیق نویسندگان به سبب شهرتی که دارند، بسی آسان‌تر خواهد شد.

### \* راهکارهای حل مسئله نام در مجلات دارای رتبه علمی

مجلات که وزارت علوم بر پایه ضوابط و مقرراتی به آنها درجه علمی-پژوهشی یا علمی-ترویجی داده است، ساختار و آیین‌نامه‌هایی برای نگارش مقالات در نظر می‌گیرند که نویسنده ملزم به رعایت آنهاست؛ به گونه‌ای که اگر مقاله‌هایی که برای نشر در این نشریات عرضه می‌شود، دارای معیارهای آنها نباشد، در نشریات مزبور منتشر نمی‌گردند. یکی از ملاک‌های متداول که مقاله‌های این نشریات باید دارا باشند، نام کامل، نشانی رایانامه، تالگان (وابستگی شغلی یا سازمانی (Affiliation)) و رتبه علمی مؤلف است. همه این اطلاعات، در مقالات نشریات علمی ثبت می‌شود. شاید از راه‌های مؤثر برای پیشگیری از مشتبه‌شدن نام پدیدآورندگان، ثبت اطلاعات تماسی و تالگانی مؤلفان باشد؛ برای نمونه، احمد احمدی که استاد فلسفه دانشگاه تهران است، با احمد احمدی که عضو هیئت علمی دانشگاه علامه طباطبایی است، تنها از طریق وابستگی‌های سازمانی و اطلاعات تماسی قابل تمییز است.

در بسیاری از پایگاه‌های مشهور، ملاک تمایز نویسندگان هم‌نام، وابستگی شغلی و سازمانی آنهاست؛ برای نمونه، پایگاه اسکوپوس که یکی از معتبرترین پایگاه‌های استنادی جهان است، از طریق وابستگی سازمانی

و تاریخ تولد افراد، میان آنها تمییز می‌دهد.

البته در نمونه‌های نادر، افزون بر نام، وابستگی سازمانی نویسندگان مقالات نیز مشابه هم می‌شود؛ در این گونه موارد، ملاکی مانند رایانامه که کاملاً یگانه است و امکان تشابه در آن نمی‌رود، می‌تواند مشکل‌گشا باشد. همچنین، دفتر تبلیغات اسلامی در پایگاه «اسلامیک داک» ([islamicdoc.org](http://islamicdoc.org)) که اصطلاح‌نامه‌های تولیدشده، نمایه‌ها و برخی متون الکترونیکی را عرضه می‌کند، صفحاتی با عنوان ویکی نیز نمایش می‌دهد که شامل ویکی: فقه، اصول فقه، علوم قرآن، اعلام و مطهر است.

بر این اساس، روش‌های گوناگونی برای استخراج نام‌ها از متون به گونه خودکار طراحی شده است که از آن جمله می‌توان به: شیوه‌های شناسایی به کمک فرهنگ اعلام و فرهنگ‌های جغرافیایی (گازترها)، شناسای نام‌ها مبتنی بر یادگیری ماشینی و شناسایی نام‌ها به شیوه قاعده‌مدار (استفاده از قوانین صرف و نحو دستور زبان) اشاره کرد.

در ویکی اعلام (<http://islamicdoc.org/wiki/index.php>) که به اعلام اشخاص اختصاص دارد، بیشتر به اعلام و مشاهیر می‌پردازد. در این مجموعه، نزدیک به ۴۰۰۰۰ نام اشخاص و مشاهیر، اعم از نام‌های مرجح و نام‌مرجح گردآوری شده که حدود ۲۷۰۰۰ نام، از اسامی مرجح است. برای برخی نام‌های مرجح، افزون بر تهیه ارجاعات، زندگی‌نامه‌ای نیز تهیه شده که اطلاعات مختصر از آنها را به کاربران ارائه می‌کند.

همچنین در قسمت دیگری از این پایگاه با عنوان «سامانه دستیار محققین»، بخشی به نام نسخه آزمایشی نمایه اعلام (<http://islamicdoc.org/electronicpublication/samaneh/aalam>) وجود دارد که در آن، کتب مرتبط با آیات قرآن، اعلام اشخاص و مکان‌ها به کاربر عرضه می‌شود.

#### پیشنهادها

مرتبط کردن اعلام به یکدیگر؛ مانند مرتبط کردن ادوار زمانی، مکان‌ها، به اشخاص یا کتب با تعیین نوع ارتباط؛ برای نمونه، علم زمانی قرن سوم را با ایران، سامانیان، ابن سینا، شفاء، اشارات، نجات، بهمنیار، همدان و خراسان بزرگ مرتبط نموده، وجه ارتباط را نیز بیان کنند؛ به این گونه که: قرن سوم هجری در ایران، سامانیان حکومت می‌کردند که ابن سینا دانشمند ایرانی، معاصر آنها می‌زیسته و آثاری چون: «شفاء»، «اشارات و تنبیهات» و «نجات» نوشته و شاگرد او، بهمنیار بود.

با این کار، از احتمال انتساب نویسنده به اثری خاص کاسته خواهد شد؛ زیرا اطلاعات همراه نویسنده، به‌ویژه زمان و مکان زندگی، اشخاص دیگر را که در زمانی پیش یا پس از آن تاریخ می‌زیسته‌اند، از دایره انتساب خارج خواهد کرد. افزون بر این، توان بازیابی دقیق را در پی خواهد داشت. این اطلاعات، زمینه استخراج دانش از متن را نیز فراهم می‌آورد.

استفاده از منابعی چون: فرهنگ اعلام فارسی، مستند مشاهیر و جای نام‌ها، برای تکمیل و تدقیق ارتباطها سودمند خواهد بود. همچنین، ثبت همه شکل‌های نگارش نام‌ها و ارجاع آنها به یکدیگر، هم در بازیابی اطلاعات و هم در نسبت دادن درست آثار به نویسندگان، سودمند می‌باشد.

#### منابع :

1. Nadeau, David, and Satoshi Sekine. "A survey of named entity recognition and classification." *Linguisticae Investigationes.*, 2007: 1-20.

۲. داماد، میرمحمدباقر. ۱۳۸۰. جزوات و موافقت. تهران: میراث مکتوب.

۳. سجادی، سیدجعفر. ۱۳۷۹. فرهنگ اصطلاحات فلسفی ملاصدر. تهران: وزارت فرهنگ و ارشاد اسلامی. ■