

آشفتگی نگارش نام پدید آورندگان ایرانی در پایگاه اطلاعاتی آی اس آی

مریم خسروی*

عضو هیئت علمی،
پژوهشگاه علوم و فناوری اطلاعات ایران

دریافت: ۱۳۸۹/۰۲/۰۸ | پذیرش: ۱۳۸۹/۰۶/۲۱

فصلنامه علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شاپا(چاپی) ۸۲۲۳-۲۲۵۱
شاپا(الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در SCOPUS و ISC
<http://jlist.irandoc.ac.ir>
ویژه نامه ذخیره، بازیابی و مدیریت اطلاعات
ص-ص ۴۵-۵۶ زمستان ۱۳۹۰
نوع مقاله: پژوهشی

*khosravi@irandoc.ac.ir

چکیده: ضبط اسامی ایرانی در منابع خارجی دچار تشتت و چندگانگی است و از قاعده خاصی پیروی نمی‌کند. پژوهش حاضر که یک مطالعه توصیفی-پیمایشی است، به منظور بررسی آشفتگی ثبت و نگارش اسامی مؤلفان ایرانی، کنکاشی در پایگاه اطلاعاتی آی اس آی، طی سال‌های ۲۰۰۷ و ۲۰۰۸ داشته است. در این راستا، فهرست کامل آثار ایرانی در بانک اطلاعاتی Web of Science، برای سال‌های ۲۰۰۷ و ۲۰۰۸ استخراج شد. از اهداف دیگر این بررسی، تعیین جامعه آماری مقالات چاپ شده ایرانیان و تعیین تعداد و درصد اسامی مؤلفان ایرانی، دستخوش آشفتگی نگارش اسامی در پایگاه اطلاعاتی آی اس آی، بود. نتایج نشان داد که نگارش اسامی مردان بیش از زنان دارای تنوع نگارش بوده است. همچنین، بیشترین میزان تنوع نگارش (در یک سوم آنها)، مربوط به کوتاه‌نویسی نام بوده است.

کلیدواژه‌ها: نام مؤلفان ایرانی، آشفتگی نگارش، نگارش نام پدید آورندگان، پایگاه اطلاعاتی آی اس آی

۱. مقدمه

نام مؤلف در تمامی پایگاه‌های اطلاعات کتابشناختی، یکی از مهمترین گزینه‌های جستجو است و اهمیت فراوانی در جستجوی منابع و یافتن آثار علمی دارد. به‌عنوان مثال، یکی از روش‌های اصلی جستجوی منابع علمی روش "جستجوی ستاره‌ها"^۱ است. به این معنی که کاوشگر اطلاعات، به جای جستجوی موضوعی، نام مؤلف برجسته‌ای را در آن حوزه موضوعی می‌شناسد و سپس، به دنبال تمامی آثار آن مؤلف می‌گردد تا آثار مرتبط با حوزه کاری خود را که توسط وی به رشته تحریر درآمده است، بیابد. جستجوی نام مؤلفان فقط برای یافتن منابع علمی به قصد مطالعه و پژوهش کاربرد ندارد، بلکه کاربردهای دیگری نیز دارد، از جمله در بررسی‌های کمی که در علم‌سنجی برای ارزیابی یک نویسنده براساس تعداد آثارش، یا میزان استناد به آثار، و محاسبه شاخص‌هایی مثل شاخص هرش^۲ (شاخص اچ) انجام می‌شود، از نام مؤلف استفاده فراوان می‌شود.

بنابراین و با توجه به اهمیت درج صحیح نام نویسندگان در پایگاه‌های اطلاعاتی و رعایت یکدستی در ثبت آنها، هرگونه تشّت و تنوع در نگارش اسامی مؤلفان می‌تواند مشکل‌آفرین باشد. این تنوع نیز در مورد اسامی نویسندگان ایرانی در پایگاه‌های اطلاعاتی آی‌اس‌آی^۳ وجود دارد و مشکلاتی را هم ایجاد نموده است. مواجهه با نامی که امکان نگارش آن به چند صورت امکان‌پذیر است، بازیابی آن مدرک را دچار اختلال می‌کند و از طرف دیگر، نمی‌توان پذیرفت که برای دستیابی به فلان مدرک، کاربر تمامی شکل‌های احتمالی نگارش نام نویسنده آن را در ذهن داشته باشد و در منابع و به همه مدخل‌ها با ضبط‌های گوناگون رجوع کند، صرف‌نظر از اینکه نباید انتظار داشت مراجعان با همه آنها آشنا باشند.

به‌عنوان مثال، نام دکتر موسوی موحّدی، یکی از پراستنادترین دانشمندان ایرانی، با ۱۴ املای مختلف در پایگاه اسکوپوس^۴ و نیز با املاهای گوناگون در پایگاه آی‌اس‌آی ثبت شده است (مهرداد ۱۳۸۸). بنابراین، جستجو برای یکی از این نام‌ها، کاوشگر را به تمامی مقالات این مؤلف هدایت نمی‌کند. حاصل این تنوع در نگارش و ثبت اسامی نیز مشکلات عدیده دیگری را ایجاد می‌کند، از جمله اشتباه در محاسبه شاخص‌های علم‌سنجی نظیر شاخص هرش و تمامی شاخص‌های دیگر که به دانشمندان مربوط می‌شود (نظیر تعداد استنادها، آثار، و غیره).

از طرف دیگر از لحاظ نگارشی، ناهماهنگی ضبط اسامی در زبان فارسی، در متون و مدخل‌های ابزارهای کتابشناختی باعث سردرگمی و آشفتگی بازار نگارش می‌شود (صدیق بهزادی ۱۳۸۱)، (به‌عنوان مثال، برای نوشتن نام مهدی به زبان لاتین یک نفر آن را به‌صورت

1. Stars Search

2. Hirsch Index

3. ISI

4. scopus

Mahdi و دیگری آن را به صورت Mehdi نگارش کرده است). همچنین تعدد نام‌ها نیز از مشکلاتی است که بر سر راه پژوهشگران قرار دارد. مواجهه با نامی که چندین قسمت دارد آنها را بی‌حوصله و سردرگم می‌کند که در نهایت، منتج به عدم توفیق در بازیابی مدرک می‌شود. مسأله قابل اشاره اینکه تنوع و آشناگی نگارش فقط مختص نام مؤلفان و برگرداندن آنها به زبان انگلیسی نیست. این امر درباره ضبط دیگر اعلام و برگرداندن آنها به زبان فارسی نیز مطرح است، آنچنان که دیگر متخصصان عنوان کرده‌اند مسأله چشمگیر دیگر، ضبط نادرست اعلام شبه قاره است. بیشتر مترجمان ایرانی در حین برگردان اسامی شبه قاره از حروف لاتین به حروف فارسی دچار خطا شده‌اند. این خطا ناشی از اختلاف‌هایی در ضوابط آوانگاری^۱ شبه قاره و ایران است. نبود آگاهی کافی در مترجمان ایرانی از صداهای مخصوص الفبای اردو / هندی و تطبیق دادن نظام آوانگاری ایران با نظام شبه قاره موجب خطا می‌شود. چند نکته به یادداشتی از نظام آوانگاری شبه قاره به شرح زیر است:

الف) برای صداهای ثقیل نظام آوانگاری شبه قاره بدین ترتیب است: به = bh؛ په = ph؛ ته = th؛ ده = dh؛ ره = rh؛ یا dh؛ که = kh؛ گه = gh. از طرفی، چون آوانگاری حروف ث th با ته و ه، خ kh با که، و غ gh با گه مشابه است، مترجمان ایرانی را دچار خلط و خطا می‌کند.

ب) برای فتحه () و الف ممدوده و الف و "ه" به‌طور معمول از یک آوا یعنی "a" استفاده می‌شود که در برگردان فارسی عدم تمیز آنها از یکدیگر موجب اشتباه شده است. گاهی فتحه به "الف" تبدیل شده است، مانند: پتیا له <PATIAL> پاتیا لا، و به فتحه گرفته شده است، مثال: مدراس <Madras> مدرس. البته گاهی در نظام دقیق آوانگاری شبه قاره برای تمیز الف ممدوده فتحه، از آوانگاری "a a" استفاده شود <P>

ج) برای کسره () از "e" استفاده می‌شود که در برگردان فارسی گاهی به "ی" تبدیل شده است، مانند: پشاور <PESHAWAR> P پیشاور.

د) برای ضمه () از آوای "U" استفاده می‌شود که مترجمان ایرانی آن را با "و" خلط کرده‌اند، مانند: ملتان <MULTAN> P مولتان (دایره‌المعارف اسلامی).

با توجه به این مقدمه و مشکل موجود در مورد نگارش لاتین اسامی نویسندگان ایرانی، این پژوهش سعی دارد تا آشناگی نگارشی اسامی پدیدآوران ایرانی را در پایگاه آی‌اس‌آی مورد مطالعه قرار دهد و میزان و چگونگی این آشناگی را مشخص نماید. در همین راستا، این پژوهش سعی دارد به سؤالات زیر پاسخ دهد:

- ۱) تعداد مقالات چاپ‌شده ایرانیان در این دو سال (۲۰۰۷ و ۲۰۰۸) چقدر بوده است؟
- ۲) تعداد مؤلفان ایرانی دارای مقاله در این پایگاه اطلاعاتی در دو سال ۲۰۰۷ و ۲۰۰۸ چگونه است؟
- ۳) پراکندگی جنسیتی مؤلفان مقالات ایرانی در این دو سال به چه صورت بوده است؟
- ۴) چه میزان از اسامی مؤلفان ایرانی دارای آشفتگی نگارش در پایگاه اطلاعاتی آی‌اس‌آی هستند؟
- ۵) چه میزان از این اسامی (که دارای آشفتگی نگارش هستند) مربوط به اسامی زنان و چه میزان آن مربوط به اسامی مردان نویسنده است؟
- ۶) نوع آشفتگی نگارش اسامی چگونه است؟

۲. پیشینه پژوهش

تا جایی که نگارنده جستجو نموده است در این مورد پژوهشی، چه در زبان فارسی و چه در زبان لاتین (از طرف زبان‌شناسان و کتابداران) انجام نشده است که از این لحاظ این پژوهش نو و بدیع است. دلیل آن نیز شاید این باشد که در زبان‌های دیگر، به‌ویژه زبان‌هایی که الفبای لاتین دارند در این باره مشکلی نیست که نیاز به پژوهش داشته باشد. همچنین امکان دارد پژوهش‌هایی، به زبان محلی و نه انگلیسی، در برخی زبان‌ها انجام و منتشر شده باشد. این مسأله، یعنی تنوع نگارش اسامی مؤلفان به زبان انگلیسی، به‌طور منطقی فقط در زبان‌هایی رخ می‌دهد که ناگزیر از آوانویسی اسامی خود به الفبای زبان انگلیسی هستند. در این میان، زبان‌های فارسی و عربی به‌احتمال، به‌دلیل استفاده خاص خود از حروف صامت و صدادار و به‌طور کل، نوع صداها و عدم وجود استاندارد در مورد آوانویسی، شرایط خاصی دارند. وجود صداها و حرکت‌های کوتاه که در زبان عربی فتحه و ضمه و کسره نامیده می‌شوند، یکی از دلایل بروز این مشکل است که این امر باعث می‌شود هر کسی بنا به ذوق و سلیقه خود دست به برگرداندن نام خود به زبان انگلیسی بزند.

در جستجوهای صورت گرفته درباره چگونگی ضبط نام افراد، صدیق بهزادی (۱۳۸۱) کتابی منتشر نموده است که در آن تلاش شده است تنوع نگارش نام افراد و نیز اعلام تاریخی و جغرافیایی در منابع فارسی نشان داده شود و نیز با ارائه قواعد مطلوب، معیارهایی برای انتخاب الگوی ضبط معرفی شود.

در جستجوی متون خارجی نیز پژوهش‌هایی درباره ابهام‌زدایی اسامی^۱ از اسامی ابهام‌دار^۲

1. Names disambiguation

2. Ambiguities names

(که تا حد کمی به این موضوع مرتبط است)، انجام شده است که به طور مختصر، به برخی از آنها اشاره می‌شود:

اسامی ابهام‌دار، مسأله‌ای چالش برانگیز است که دارای سابقه تاریخی طولانی است. در زمان مدیریت اطلاعات دنیای واقعی، با مسأله‌ای به نام ابهام‌زدایی اسامی مواجه هستیم. ابهام اسامی ممکن است به دو دلیل رخ دهد: شخصی چند نام داشته باشد و یا چند نفر یک نام داشته باشند (Masada and Adachi 2007).

در حیطه کتابخانه‌های دیجیتال، پیدا کردن اسامی ابهام‌دار در اسنادهای کتابشناختی، امری عادی است. ابهام ممکن است به علت وجود چندین مؤلف با اسامی مشابه و یا اسامی متفاوت برای یک مؤلف پیش بیاید. به طور اخص، دو مسأله اسنادهای مرکب و اسنادهای مجزا ممکن است در این رابطه صورت بگیرد. اسنادهای مرکب به این صورت است که چند مؤلف، یک نام و یا نامی که در هنگام تلفظ شبیه به هم باشند، دارند. برای مثال، دی پریرا^۱ امکان دارد به عنوان دنیلسون پریرا^۲ ارجاع شود که در این صورت دو شخص متفاوت هستند. اسنادهای مجزا نیز هنگامی اتفاق می‌افتد که اسنادهای یک مؤلف با اسامی متفاوت آورده می‌شود. برای مثال، دنیلسون آلوس پریرا^۳ که ممکن است به صورت مختصر "دنیلسون پریرا" یا "دی پریرا" و یا "دی ای پریرا" و یا با تلفظ غلط مانند "دنیلسون فریرا" باشد (Pereira et al. 2009).

ابهام اسامی ممکن است بر کیفیت گردآوری اطلاعات علمی تأثیر بگذارد و یا می‌تواند باعث کاهش عملکرد بازیابی اطلاعات و جستجو در وب باشد؛ حتی ممکن است باعث شناسایی و استناد اعتباری نادرست به مؤلفان شود. این مسأله می‌تواند چندین علت داشته باشد: به دلیل گوناگونی اسامی، اسامی همانند، تلفظ غلط اسامی، و یا نام‌های مستعار. دو نوع ابهام اسامی را در مقالات پژوهشی و کتابشناختی‌ها (اسنادها) می‌توان مشاهده نمود. نوع اول اینکه یک مؤلف چندین نام داشته باشد، برای مثال: مؤلف "دیوید اس جانسون" یا "دی اس جانسون"، و یا با املا نادرستی مانند "داوید جانسون". دومین نوع اینکه چندین مؤلف دارای یک نام مشابه باشند، برای مثال: "دی جانسون" را ممکن است به صورت "دیوید بی جانسون" از دانشگاه رایس ارجاع داده شود و یا "دیوید ای جانسون" از آزمایشگاه تحقیقی AT & T و یا "دیوید ای جانسون" از دانشگاه یوتا باشد (Han, Zha, and Giles 2005).

۳. روش پژوهش، جامعه آماری، و روش گردآوری داده‌ها

پژوهش حاضر یک مطالعه توصیفی-پیمایشی است. جامعه پژوهش شامل همه

1. D. Pereira

2. Denilson Pereira

3. Denilson Alves Pereira

نویسندگان و مؤلفان ایرانی است که مقاله یا مقاله‌های خود را در سال‌های ۲۰۰۷ و ۲۰۰۸ در پایگاه اطلاعات آی‌اس‌آی به‌چاپ رسانده‌اند. تعداد کل آنها پس از اعمال محدودیت مورد نظر (فقط مقاله و افراد دارای نام کوچک)، طبق رکوردهای ثبت‌شده در این پایگاه اطلاعاتی و در این برهه زمانی، ۳۹۹۹۸ رکورد بود. در این پژوهش، کل جامعه آماری مورد مطالعه قرار گرفت. لازم به اشاره است که افرادی هم که نام کوچک آنها به‌صورت کوتاه و مختصر ثبت شده بود، ولی امکان جستجو و تعیین جنسیت آنها از طریق اینترنت میسر بود، در این جامعه قرار گرفتند.

برای گردآوری اطلاعات، ابتدا فهرست کامل آثار ایرانیان در بانک اطلاعاتی Web of Science برای سال‌های ۲۰۰۷ و ۲۰۰۸ به‌صورت (tagged format (marked list) در قالب دسته‌های پانصدتایی استخراج شد. سپس، این رکوردها در هم‌کرد شد و با استفاده از نرم‌افزار نوشته‌شده توسط لوئت لیدسدورف^۱ به تعدادی فایل با فرمت بانک اطلاعاتی تبدیل گردید. پس از آن، اسامی مورد نظر، در نرم‌افزار Access 2007 وارد گردید. پس از به‌نظم الفبای درآوردن اسامی، اطلاعات آنها در چندین مرحله تکمیل شد. در مرحله اول، در صورتی که در ثبت و نگارش نام و نام خانوادگی آنها هرگونه (برای افراد دارای چندین مقاله و یا دارای غلط تایپی) آشفتگی (تفاوت) مشاهده می‌شد، حرف y و چنانچه یکسان بود حرف n جلوی نام آنها درج شد. در مرحله بعد، برای هر یک از آنها تعیین جنسیت (به‌وسیله نام کوچک) انجام شد. لازم به اشاره است که در این مرحله، گاهی به‌علت تشابه نوشتاری انگلیسی برخی اسامی هم‌آوا (مانند شهین و شاهین؛ و یا ثمن و سامان) که صورت واج‌نگاری مشابه دارند، از جستجو در اینترنت استفاده شد. در مرحله آخر، برای هر یک از آنها کد مشخص گردید. در این مرحله، در صورتی که نام فرد مورد بررسی (دارای چندین مقاله) دارای آشفتگی نگارش بود، کد تعریفی برای نام دارای تنوع نیز تکراری تعریف می‌شد.

برای بررسی آشفتگی نوع نگارش، ۱۰٪ از اسامی دارای گوناگونی نگارش به‌عنوان نمونه، مورد مشاهده قرار گرفتند تا نوع آشفتگی نگارش اسامی آنها مشخص شود.

در پایان، اطلاعات جمع‌آوری‌شده، با استفاده از روش‌های آماری توصیفی چون توزیع فراوانی و درصد، مورد تجزیه و تحلیل قرار گرفتند.

۴. نتایج

۴-۱. بررسی وضعیت مقالات چاپ‌شده

جامعه آماری پژوهش در جدول ۱ مشاهده می‌شود:

1. Loet Leysdorff

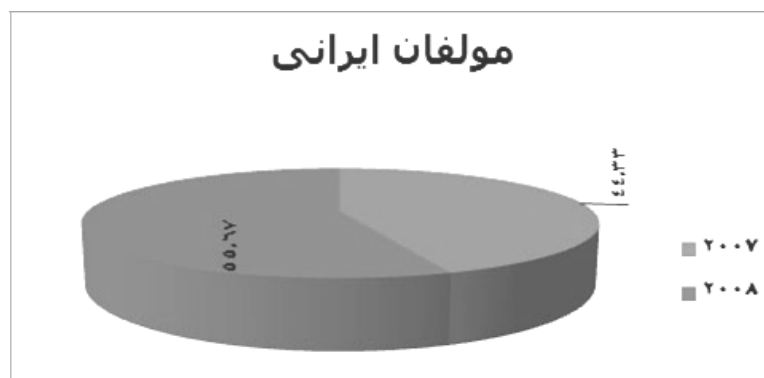
جدول ۱. جامعه آماری بر حسب جنسیت

جنسیت	مقالات چاپ شده	
	تعداد	درصد
زن	۸۷۴۲	٪۲۱/۸۶
مرد	۳۱۲۵۶	٪۷۸/۱۴
جمع	۳۹۹۹۸	۱۰۰

آن چنان که در جدول ۱ و نمودار ۱ نیز نشان داده می شود، از تعداد ۳۹۹۹۸ مقاله چاپ شده در این پایگاه اطلاعاتی، مقالاتی که توسط مردان در این پایگاه اطلاعاتی به چاپ رسیده است بیش از ۳/۵ برابر زنان است. (البته لازم به اشاره است که این نتایج با توجه به محدودیت های پژوهش، به لحاظ اینکه برای برخی از اسامی مؤلفان شناسایی جنسیت آنها مقدور نبود، احتمال خطا دارد).

جدول ۲. انتشار مقاله توسط ایرانیان

سال انتشار	جمع	درصد
۲۰۰۷	۱۷۷۳۰	۴۴/۳۳
۲۰۰۸	۲۲۲۶۸	۵۵/۶۷
جمع	۳۹۹۹۸	۱۰۰



نمودار ۱. مقایسه سال های مورد مطالعه و چاپ مقاله توسط ایرانیان

۲-۴. بررسی جنسیت مؤلفان و سال انتشار

جدول ۳ تعداد مقالات چاپ شده توسط ایرانیان را در این دو سال، به تفکیک جنسیت آنها

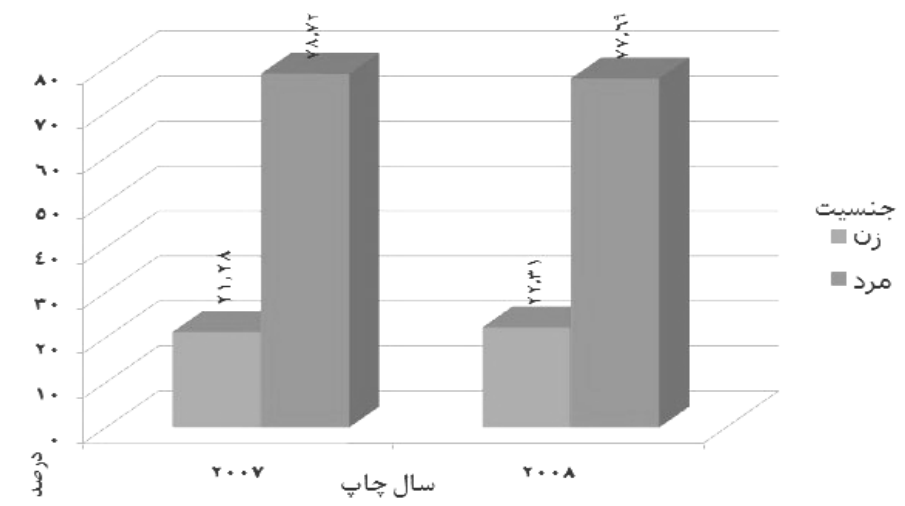
نشان می دهد:

جدول ۳. تعداد مقالات چاپ‌شده

جمع		سال چاپ				جنسیت
		۲۰۰۸		۲۰۰۷		
درصد	تعداد	درصد	تعداد	درصد	تعداد	
۲۱/۸۶	۸۷۴۲	۲۲/۳۱	۴۹۶۹	۲۱/۲۸	۳۷۷۳	زن
۷۸/۱۴	۳۱۲۵۶	۷۷/۶۹	۱۷۲۹۹	۷۸/۷۲	۱۳۹۵۷	مرد
۱۰۰	۳۹۹۹۸	۱۰۰	۲۲۲۶۸	۱۰۰	۱۷۷۳۰	جمع

براساس جدول ۳ و نیز نمودار ۲، در سال ۲۰۰۷ تعداد ۱۷۷۳۰ مقاله توسط ایرانیان در این پایگاه اطلاعاتی به چاپ رسیده است، درحالی‌که این امر در سال ۲۰۰۸ افزایش داشته است (۲۲۲۶۸ مقاله).

از تعداد ۱۷۷۳۰ مقاله چاپ‌شده در سال ۲۰۰۷، تعداد ۳۷۷۳ مقاله توسط زنان نگارش شده و سهم مردان در سال ۲۰۰۷ تا حد ۳/۵ برابر زنان بوده است. در سال ۲۰۰۸ نیز ۲۲۲۶۸ مقاله به چاپ رسیده که ۴۹۶۹ مقاله توسط زنان و ۱۷۲۹۹ مقاله دیگر مربوط به مؤلفان مرد بوده است. در کل، از تعداد ۳۹۹۹۸ مقاله به چاپ رسیده در این دو سال، ۸۷۴۲ مقاله آن توسط زنان (۲۱/۸۶٪) و ۳۱۲۵۶ مقاله توسط مردان (۷۸/۱۴٪) نگارش شده است.



نمودار ۲. مقایسه سال چاپ مقاله و جنسیت

۳-۴. بررسی آشفتگی نگارش اسامی

جدول ۴ تعداد مؤلفان ایرانی دارای مقاله در این دو سال و تعداد اسامی دارای آشفتگی نگارش را نشان می‌دهد (برخی از مؤلفان دارای چند مقاله بوده‌اند):

جدول ۴. آشفتگی نگارش اسامی مؤلفان

نگارش اسامی		جنسیت			
		مرد		زن	
جمع	تعداد	درصد	تعداد	درصد	تعداد
آشفتگی نگارش	۲۲۸۱	۱۵/۱۲	۱۸۸۴	۸/۳۲	۳۹۷
عدم آشفتگی نگارش	۱۴۹۵۲	۸۴/۸۸	۱۰۵۷۵	۹۱/۶۸	۴۳۷۷
جمع	۱۷۲۳۳	۱۰۰	۱۲۴۵۹	۱۰۰	۴۷۷۴

همان‌طور که مشاهده می‌شود در کل، ۱۷۲۳۳ مؤلف در این دو سال در این پایگاه اطلاعاتی مقالاتی را به چاپ رسانده‌اند که از این تعداد، ۲۲۸۱ (۱۳/۲۴٪) مؤلف در نگارش نام و یا نام خانوادگی خود دارای آشفتگی نگارش بوده و ۱۴۹۵۲ (۸۶/۷۶٪) آنها همیشه به یک صورت عمل نموده‌اند. در مقایسه آشفتگی نگارش، اسامی زنان (۱۵/۱۲٪) بسیار کمتر از مردان (۸۴/۸۸٪) بوده است.

۴-۴. بررسی نوع آشفتگی نگارش اسامی

برای بررسی نوع تنوع نگارش اسامی، تعداد ۲۲۰ مورد از نگارش اسامی دارای تنوع، بررسی شد. گوناگونی مشاهده‌شده در جدول ۵ قابل رؤیت است (لازم به اشاره است که برخی از اسامی دارای چندین تنوع نگارش نام بوده‌اند):

جدول ۵. بررسی تنوع نگارش

درصد	تعداد	کوتاه‌نویسی	شناسایی تنوع
۲۹/۵۲	۹۸	غلط تایپی و حروف صامت	
۲۸/۹۲	۹۶	دوقسمتی بودن نام کوچک	
۱۷/۱۷	۵۷	عدم رعایت فاصله	
۱۱/۷۵	۳۹	حروف صدادر	
۱۰/۸۴	۳۶	دوقسمتی بودن نام خانوادگی	
۱/۸۰	۶	جمع	
۱۰۰	۳۳۲		

همان‌طور که ملاحظه می‌شود، ۲۹/۵۲٪ از گوناگونی‌های مشاهده‌شده مربوط به اختصار و کوتاه‌نویسی است. پس از آن، غلط‌های تایپی به میزان ۲۸/۹۲٪ قرار دارد. این درحالی است که دو قسمتی بودن نام کوچک نیز می‌تواند باعث ۱۷/۱۷٪ از گوناگونی‌ها باشد. عدم رعایت فاصله و یا گذاشتن خط فاصله بین دو قسمت نام کوچک و یا نام خانوادگی نیز باعث ۱۱/۷۵٪ از تنوع نگارش‌ها بوده است. حروف صدادار /e//æ//o//:u//:i//:a/ نیز ۱۰/۸۴٪ از تنوع‌های نگارش و در انتها، دو قسمتی بودن نام خانوادگی باعث ۱/۸۰٪ از تنوع نگارش‌ها شده‌اند.

۴-۵. انواع آشفتگی‌های نگارش مشاهده‌شده

در ادامه، چند صورت از آشفتگی نگارش اسامی مشاهده‌شده، آورده می‌شود:
کوتاه‌نویسی شامل نوشتن فقط حروف اول نام، به‌عنوان مثال:

Abadpour, A
Abadpour, Arash
Aalami, M
Aalami, Mehran

غلط تایپی و حروف صامت شامل:

Bazgirb, Ayoob
Bazgir, Ayoob
Bazrafshan, Afireza

دو قسمتی بودن نام مانند:

Abbasi, Mohammad Reza
Abbasi, Mohamad R.

دو قسمتی بودن نام خانوادگی همچون:

Bozorgmehry, Bozarjomehry
Bozorgmehry, B.
Forouzandeh-Moghadam, Mehdi
Forouzandeh, M.

عدم رعایت فاصله و گذاشتن خط فاصله بین دو قسمت نام، شامل:

Barqi, Mohamadali
Abbasi, Mohamad- Hassan

حروف صدادار مانند:

Abbasi, Massoumeh
Abbasi, Massuomah
Abbasi, Shahriar
Abbasi, Shahryar

لازم به اشاره است که اینچنین تنوع، درباره نام خانوادگی نیز صادق بود.

Ebrahimnezhad, Yahya
Ebrahim-Nezhad, Yahya
Ebrahimnezhad, Y

و یا؛

Bezzaatpour, Abolfazl
Bezaatpour, Abolfazl

۵. بحث و نتیجه‌گیری

نام‌ها کلماتی هستند که از هزاران سال پیش فقط برای تشخیص و تمیز افراد از یکدیگر وضع شده‌اند. چگونگی ضبط نام افراد در نوشته‌ها، به‌خصوص در فهرست منابع و مآخذ و در کتاب‌های مرجعی نظیر دانشنامه‌ها از مسائلی است که در عرصه ترجمه و تألیف مطرح می‌شود. از آنجا که ضبط اسامی ایرانی در منابع انگلیسی دچار تشبّه و آشفته‌گی است و چون در این حوزه از قواعدی پیروی نشده است، هر کس به‌سلیقه و ذوق خود، ضبط و صورتی را انتخاب کرده، در نتیجه یک اسم به دو یا چند شکل نوشته شده است (صدیق بهزادی ۱۳۸۱). این امر باعث شده است که بازبایی تمام مقالاتی که از طرف یک نویسنده به چاپ رسیده است، قابل بازبایی نباشد و یا به‌علت تشابه اسامی، چندین مقاله به یک مؤلف استناد داده شود. در همین راستا، این پژوهش به بررسی نحوه ثبت اسامی پدیدآورندگان ایرانی در بانک اطلاعاتی آی‌اس‌آی در سال‌های ۲۰۰۷ و ۲۰۰۸ پرداخته است تا مشخص کند که چه درصدی از اسامی مؤلفان ایرانی در این پایگاه دارای آشفته‌گی نگارش هستند. نتایج نشان داد که از ۳۹۹۹۸ مقاله چاپ‌شده ایرانیان در این پایگاه اطلاعاتی در این دو سال، که متعلق به ۱۷۲۳۳ مؤلف است (البته ممکن است از این تعداد بیشتر باشد، ولی به‌علت تشابه نام، یک نفر در نظر گرفته شده باشد)، تعداد ۱۴۵۹۲ نفر از مؤلفان برای نگارش نام خود همیشه به یک صورت عمل نموده و ۲۲۸۱ نام دارای آشفته‌گی نگارش بوده است؛ همچنین، ۳۹۷ نفر آنان زن و ۱۸۸۴ نفر مرد است.

منظور از آشفته‌گی نگارش نام در این پژوهش، هرگونه اختلاف نوشتاری است، از جمله کوتاه‌نویسی نام و یا نام خانوادگی (برای مثال، برای نوشتن نام Ali Reza یک‌بار آن را کامل و بار دیگر یک قسمت از آن مثل A. Reza و یا به‌صورت Ali. R. نوشته شده است)، غلط‌تایی، خط فاصله‌گذاردن بین دو قسمت نام و یا نام خانوادگی، فاصله‌قرار ندادن بین دو قسمت نام و یا نام خانوادگی، متفاوت نوشتن حروف صدادار (به‌عنوان مثال، برای نوشتن نام محمود، مسعود و یا سعید ... که به صورت‌های Saied, Saeed, / Masoud, Masood / Mahmoud, Mahmud, Mahmuud سعید عمل شده است). این آشفته‌گی نگارش در بسیاری از اسامی و نیز نام خانوادگی (به‌عنوان مثال، برای نوشتن نام خانوادگی صفایی، موسوی، اکبرنژاد به چند صورت زیر عمل شده است: Safaii, Safaie, Safaiy, Safaee Akbarnejad, Akbarnejhad Mosavi, Moosavi, Mousavi, Mousavy)

در کدگذاری و همچنین نظم‌الفبایی آنها، مشکلات عدیده‌ای را ایجاد می‌کند که باعث شد چندین بار کدگذاری آنها تعویض شد.

در بررسی‌ای که به منظور نوع آشفتگی نگارش اسامی به عمل آمد، تعداد ۲۲۰ مورد از نام‌هایی که به عنوان دارای آشفتگی نگارش شناخته شده بودند، مورد بازبینی قرار گرفتند و شش نوع آشفتگی نگارش رایج در بین آنها یافت شد، اما بیشترین نوع آن مربوط به اختصار و کوتاه‌نویسی (۲۹/۵۲٪) بود.

۶. پیشنهادها

- با توجه به ناهماهنگی ضبط اعلام در منابع و ابزارهای کتابشناختی:
- نیاز به تدوین شیوه‌نامه‌ای واحد برای برطرف کردن تشابه موجود، از سوی وزارتخانه متبوع که در آن مفاهیم ظریف و دقیق surname, initial name, middle name, family به طور دقیق تعریف شده است و از یکدیگر افتراق می‌یابند، در ابزارهای کتابشناختی و متون تدوین یافته احساس می‌شود.
 - فهرست مستند اسامی برای یکدست شدن نام و نام خانوادگی پدیدآورندگان، به منظور حصول اطمینان از یکدستی کاربرد آنها تهیه شود.
 - همچنین، بهتر است شیوه‌نامه‌ای ملی برای آوانویسی اسامی ایرانی به انگلیسی تهیه شود تا در موارد مختلف برای رعایت یکدستی از آن استفاده شود. به عنوان مثال، برگرداندن نام افراد به انگلیسی برای درج در پاسپورت و نیز درج اسامی مؤلفان ایرانی به انگلیسی در آثار آنان.

۷. منابع

- صدیق‌بهبزادی، ماندانا. ۱۳۸۱. شیوه‌نامه ضبط اعلام. تهران: نشر آثار.
- مهرداد، جعفر. ۱۳۸۸. سخنرانی در باب وضعیت تولید علم. تهران: دانشگاه تهران
- نوشاهی، عارف. نگاهی به دایره‌المعارف‌های اسلامی. مجله میراث آئینه ۲۲. <http://www.hawzah.net/Hawzah/Articles/Articles.aspx> (دسترسی در ۸۹/۲/۱).
- Han, H., Zha, H., and L. C. Giles. 2005. Name disambiguation in author citation using a k-way spectral clustering method. <http://www.library.umass.edu/reference/glossary.html> (accessed 9 Sep. 2010).
- Masada, T., and J. Adachi. 2007. Citation data clustering for author name disambiguation. <http://www.cis.nagasaki-u.ac.jp/~masada/masadaINFOSCALE2007.pdf> (accessed 9 Sep. 2010).
- Pereira, D., A. Ribeiro-Neto, N. Ziviani, F. Laender, H. Alberto, A. Ferreira, and M. A. Goncalves., 2009. Using web information for author name disambiguation. <http://www.arnetminer.org/viewpub.do?pid=1212357> (accessed 9 Sep. 2010).



Study of Spelling Negligence of Iranian Authors' Names in "ISI" Database

Maryam Khosravi*
Faculty Member in IRANDOC

Iranian Journal of
**Information
Processing &
Management**

Abstract: The recording of Persian names suffers from negligence and no rule is being followed in this regard. This paper was a descriptive-quantitative study, it also examined negligence recording of names of Iranian authors in the ISI database. In line with the study goals, the complete list of papers by Iranians during 2007 and 2008 in the Web of Science Database were extracted. Of other purpose of this research was to determine their gender and their way of spelling (consistently or negligently) were surveyed and each of the authors was dedicated a record. The results showed men's names were more negligently recorded compared with that of women.

Keywords: Names of Iranian authors, spelling negligently, spelling of authors' names, ISI database

Iranian Research Institute
For Science and Technology
ISSN 2251-8223
eISSN 2251-8231
Indexed in LISA, SCOPUS & ISC
special issue: on Information Storage,
retrieval and Management (winter 2012)

*Corresponding author: khosravi@irandoc.ac.ir