

The Challenges of Transliteration of Persian Names to English, and their Impact on the Recall of Web of Science's Retrieved Results

Mahsa Kaveh

M.A. in Knowledge and Information Science; Shiraz University; Shiraz, Iran Email: mahsa.kaveh1992@gmail.com

Mahdieh Mirzabeigi*

PhD in Knowledge and Information Science; Associate Professor; Shiraz University; Shiraz, Iran Email: mmirzabeigi@gmail.com

Hajar Sotudeh

PhD in Knowledge and Information Science; Associate Professor; Shiraz University; Shiraz, Iran Email: sotudeh@shirazu.ac.ir

Amirsaeid Moloodi

PhD in Linguistics; Assistant Professor; Shiraz University; Shiraz, Iran Email: amirsaeid.moloodi@gmail.com

Received: 13, May 2019 Accepted: 07, Mar. 2020

Abstract: The present study investigated the challenges of transliteration of Persian names to English ones, and the impact of these challenges on the recall of web of science's retrieved results. This study is applied in terms of objectives and quantitative (i.e., survey) in terms of data collection procedure. The population included the names of Iranian authors who had published an English article in Web of Science between 2010 and 2017. First of all, 3110873 names were retrieved. It was reduced to 11, 242 after homogenization process. Data was collected through Bibexcel and Excel software and Web of Science database. The findings showed that there were three types of challenges (i.e., vowels, consonants, and omitted or double consonants) in transliteration of Persian names to English ones. They resulted in non-homogenous and different written forms. In addition, the results of Chi-square test showed that the frequency of all forms of the mentioned challenges had a significant effect on the recall of the retrieved results. Considering the importance of retrieving the names in retrieval studies as well as research studies' evaluation, some solutions such as using Boolean operators and wildcards between the prevailing typographical forms, creating Persian soundex, and equipping the database with the thesaurus are suggested.

* Corresponding Author

Iranian Journal of
Information
Processing and
Management

Iranian Research Institute
for Information Science and Technology
(IranDoc)

ISSN 2251-8223

eISSN 2251-8231

Indexed by SCOPUS, ISC, & LISTA

Vol. 35 | No. 4 | pp. 1065-1094

Summer 2020



Keywords: Recall, Information Retrieval, Studies' Evaluation, Transliteration of Persian Names to English, Names Transliteration Challenges

شناسایی چالش‌های برگردان اسامی فارسی به انگلیسی در پایگاه وب آوساینس

مهسا کاوه

کارشناسی ارشد علم اطلاعات و دانش‌شناسی؛
دانشگاه شیراز؛ شیراز، ایران؛
mahsa.kaveh1992@gmail.com

مهدیه میرزاییگی

دکتری علم اطلاعات و دانش‌شناسی؛ دانشیار؛
دانشگاه شیراز؛ شیراز، ایران؛
mmirzabeigi@gmail.com پدیدآور رابط

هاجر ستوده

دکتری علم اطلاعات و دانش‌شناسی؛ دانشیار؛
دانشگاه شیراز؛ شیراز، ایران sotudeh@shirazu.ac.ir

امیرسعید مولودی

دکتری زبان‌شناسی؛ استادیار؛ بخش زبان‌های خارجی و
زبان‌شناسی؛ دانشگاه شیراز؛ شیراز، ایران؛
amirsaeid.moloodi@gmail.com



دریافت: ۱۳۹۸/۰۲/۲۳ | پدیش: ۱۳۹۸/۱۲/۱۷ | مقاله برای اصلاح به مدت ۱۰۹ روز نزد پدیدآوران بوده است.

چکیده: هدف پژوهش حاضر، شناخت چالش‌های برگردان نام‌های فارسی به انگلیسی و فراوانی این چالش‌ها در پایگاه «وب آوساینس» است. این پژوهش به لحاظ هدف، کاربردی و از نظر نحوه گردآوری داده‌ها، کمی از نوع پیمایشی است. جامعه پژوهش شامل نام‌های نویسندگان ایرانی در پایگاه «وب آوساینس» است که در بازه زمانی هفت‌ساله ۲۰۱۰ تا ۲۰۱۷ مقاله انگلیسی نگاشته‌اند. تعداد ۳۱۱۰۸۷۳ نام در ابتدا بازیابی شد که پس از پالایش و یکدست‌سازی اسامی به ۱۱۲۴۲ نام تقلیل یافت. به‌منظور پالایش، آماده‌سازی و تحلیل داده‌ها از نرم‌افزار «یب‌اکسل»، «اکسل» و پایگاه «وب آوساینس» استفاده شد. یافته‌های پژوهش نشان داد که در برگردان نام‌ها از فارسی به انگلیسی چهار گروه چالش‌های مربوط به «همخوان‌ها»، چالش‌های مربوط به «واکه‌ها»، چالش‌های مربوط به «حذف یا تکرار حروف»، و چالش‌های «تلفظی» وجود دارد. در دسته اول، چالش‌های همخوان‌ها، نویسندگان و نگارندگان در برگردان نام‌ها در یازده حرف (۸ واج)، شامل حروف ق، غ، ژ، ع، و، ک، ج، ک، ث، س، ص، صورت‌های مختلف نگارشی داشته‌اند. در دسته دوم، چالش واکه‌ها، صورت‌های نگارشی مختلفی در برگردان واکه‌های

فصلنامه | علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
(ایرانداک)

شاپا (چاپی) ۲۲۵۱-۸۲۲۳

شاپا (الکترونیکی) ۲۲۵۱-۸۲۳۱

نماینده در SCOPUS، LISTA، ISC، و

jipm.irandoc.ac.ir

دوره ۳۵ | شماره ۴ | صص ۱۰۶۵-۱۰۹۴

تابستان ۱۳۹۹



(/i/، /u/، /e/ و /o/) و پایانه (ـه) از فارسی به انگلیسی مشاهده شد. در دسته سوم، وجود حروف مشدد و حروف (ال) در برخی نام‌ها موجب به وجود آمدن چندگانگی در نگارش نام‌ها شده است. و در نهایت، در در دسته آخر، آوای ا و او، و همچنین حذف برخی حروف به علت تفاوت در تلفظ نام‌ها در گویش‌ها و زبان‌های مختلف دچار چندگانگی شده‌اند. با توجه به اهمیت بازیابی اسامی در جست‌وجوی نویسندگان و همچنین در ارزیابی پژوهش، راهکارهایی همچون استفاده از عملگرهای بولی و نویسه جایگزین بین صورت‌های نگارشی غالب، ایجاد ساندکس زبان فارسی، مجهز کردن پایگاه به ابزارهای دانشی ویژه نام جهت یکدستی در برگردان نام‌ها ضرورت دارد.

کلیدواژه‌ها: بازیابی اطلاعات، پایگاه وب آوساینس، برگردان انگلیسی به فارسی، چالش‌های برگردان نام‌ها

۱. مقدمه

نشر الکترونیکی، افزایش روزافزون مقالات، و ورود اطلاعات آن‌ها به پایگاه‌های فراداده بدون بررسی و نظارت لازم بر اطلاعات کتابشناختی به دشواری‌های بازیابی اطلاعات بیش از پیش دامن می‌زند. ناشران بر صحت ورود اطلاعات منابع حساسیت و نظارت چندانی ندارند و اسامی اغلب با نام کامل و یا حروف آغازین نوشته می‌شوند. همچنین، تنوع نگارش نام به وسیله نویسندگان به دلیل ویژگی‌های زبانی، فرهنگی و اشتباهات تایپی جزء امور متداول است. در عین حال، نویسندگان از جمله مهم‌ترین عوامل تولید اطلاعات در جوامع علمی به‌شمار می‌آیند. شناسایی نویسنده مهم‌ترین عنصر تحلیل کتابشناختی است. همچنین، شناسایی نویسنده، به دانشمندان حوزه علم‌سنجی اجازه می‌دهد تا پژوهشگران را در سطح فردی مورد مطالعه قرار دهند. داده‌ها در سطح فردی به محققان کمک می‌کند تا رشد و تحرک داخلی و بین‌المللی هر پژوهشگر را ردیابی کنند. این امر می‌تواند به توسعه سیاست‌های علم و فناوری مبتنی بر منابع انسانی کمک کند (Kawashima & Tomizawa 2015).

علی‌رغم اهمیت دسترسی صحیح به اطلاعات مربوط به یک نویسنده، این امر مشکل بوده و به‌سادگی نمی‌توان با استفاده از اسمی که در اختیار است تمامی تولیدات علمی آن فرد را بازیابی کرد. افزایش دسترسی به منابع اطلاعاتی ناهمگون و تنوع برگردان اسامی به دلیل ویژگی‌های زبانی و فرهنگی از سویی و ناتوانی پایگاه‌ها در تشخیص ارتباط و پیوند میان آن‌ها از سوی دیگر، باعث گسیختگی و نبود انسجام در بازیابی اسامی افراد

شده است (صادقی گورجی و همکاران ۱۳۹۴). برگردان اسامی خاص تنها به نام نویسنده مقاله محدود نمی‌شود، بلکه در فرایند استناددهی نیز مشاهده می‌گردد. برگردان، فرایندی است که در آن یک رشته ورودی از حروف به رشته‌ای دیگر از حروف تبدیل می‌شود و معمولاً بر اساس آوای کلمه اصلی انجام می‌شود (Josan & Lehal 2010). برگردان اسامی معمولاً تابع دو حالت است: برگردان روبه‌جلو و برگردان روبه‌عقب. در برگردان روبه‌جلو بسیاری از تغییرات در نام پذیرفته شده است. این مسئله به‌ویژه زمانی اتفاق می‌افتد که برگردان بین دو زبان با ناسازگاری‌های آوایی زیاد مانند فارسی یا عربی به انگلیسی صورت می‌پذیرد. برای مثال، نام «یاسر» می‌تواند به شیوه‌های مختلفی مانند Yasir, Yassir, Yaser, Yasser برگردان شود. در برگردان روبه‌عقب، فرایند برگردان به صورت معکوس اتفاق می‌افتد؛ زیرا اصل نام به دست می‌آید. در این حالت، تنها یک برگردان مورد قبول است؛ برای مثال، برای نام «روبرت» یا «روپیرت» تنها صورت نگارشی Robert برگردان می‌شود (Al-Onaizan & Knight 2002).

همچنین، برگردان اسامی از زبان مبدأ به زبان مقصد می‌تواند تابع چندین تغییر باشد. برای برگردان اسامی خاص از زبان مبدأ به زبان مقصد، این اسامی رونویسی می‌شوند؛ مانند New York (نیویورک)؛ برخی به زبان مقصد برگردان می‌شوند و قوانین واجی آن‌ها مطابق آن زبان تغییر می‌کند؛ مانند Spain (اسپانیا)، در برخی برگردان‌ها ترجمه صورت می‌پذیرد؛ مانند Queen Elizabeth (ملکه الیزابت)؛ گروه بعدی اسامی خاص با معادل کلمات در زبان محلی جایگزین می‌شوند؛ مانند Greece (یونان)؛ و در گروه آخر، اسامی زبان مبدأ به صورت کلمه‌به‌کلمه به زبان مقصد ترجمه می‌شوند؛ مانند Islamic Republic of Iran (جمهوری اسلامی ایران) (Falih 2009).

در برگردان اسامی دو نوع ابهام برای اسامی وجود دارد که شامل اسامی یک نویسنده با شکل‌های نگارشی مختلف و اسامی چند نویسنده با یک شکل نگارشی است (Kim & Kim 2019). در رابطه با اسامی یک نویسنده با شکل‌های نگارشی مختلف، بازیابی جامع اطلاعات بدون شناخت شکل‌های نگارشی مختلف اسم نویسنده مشکل است. از سوی دیگر، اسامی چند نویسنده با یک شکل نگارشی منجر به بازیابی اطلاعات نامرتب از نویسندگان دیگر با همان شکل نگارشی نام نویسنده مورد نظر می‌شود. علی‌رغم اهمیت یکدستی در انتخاب اسامی پدیدآورندگان، این امر چندان مورد توجه قرار نمی‌گیرد و صور مختلف نگارشی برای اسامی یک پدیدآور مشاهده می‌شود.

تاکنون شیوه‌های مختلف برگردان اسامی فارسی به انگلیسی بررسی نشده است. بر این اساس نه تنها تصویری از وضعیت برگردان اسامی وجود ندارد، بلکه چالش‌هایی که باعث تنوع احتمالی برگردان یک اسم خاص می‌شود نیز شناسایی نشده است. پژوهش حاضر تلاش دارد که ضمن شناسایی چالش‌های برگردان اسامی فارسی به انگلیسی، فراوانی این چالش‌ها را شناسایی کند تا از این رهگذر ضمن شناخت وضعیت موجود برگردان اسامی پژوهشگران ایرانی از فارسی به انگلیسی و شناسایی رفتارهای نگارشی غالب نویسندگان و پژوهشگران، راهکارهایی در جهت بهبود چالش‌های احتمالی ارائه گردد. با توجه به مسئله گفته شده هدف کلی پژوهش شناسایی چالش‌های برگردان نام‌های فارسی به انگلیسی نویسندگان ایرانی در پایگاه «وب‌آوساینس» و فراوانی هر یک از چالش‌های شناسایی شده است.

۲. پیشینه پژوهش

چالش‌های موجود در زبان فارسی در سه سطح معنایی، نحوی و ریخت‌شناسی تقسیم‌بندی می‌شوند و چالش‌های برگردان نام‌ها را معمولاً در سطح معنایی قرار می‌دهند. عمده پژوهش‌های انجام شده بر چالش‌های ریخت‌شناسی زبان فارسی و راهبردهای برگردان تمرکز داشته‌اند و کمتر پژوهشی به بررسی چالش‌های مرتبط با برگردان اسامی فارسی به انگلیسی اشاره داشته است (Fumani, Goltaji & Parto (2013 و «گلتاجی و عباس‌پور» (۱۳۹۶) در پژوهش‌های خود به بررسی برگردان نام دانشگاه‌ها و مراکز پژوهشی پرداخته‌اند. علاوه بر پژوهش‌های پیش گفته، توجه به موانع به اشتراک گذاری مستندات اسامی اشخاص در «فهرست مستند مجازی بین‌المللی»^۱ (ریاحی‌نیا و نیک‌نیا ۱۳۹۵)، عدم توانایی موتور جست‌وجوی «گوگل پژوهشگر» در بازیابی اطلاعات و انتشارات نویسندگان پُرستناد بر اساس هر نوع نگارش نام (صادقی گورجی و همکاران ۱۳۹۴)، موفقیت استفاده از روش رخدادمحور برای تعیین گونه غالب املائی اسامی نویسندگان خارجی (فلاحتی قدیمی فومنی ۱۳۹۲)، ارائه الگوریتم جدید برای برگردان انگلیسی به فارسی (Karimi, Scholer & Turpin 2007) و تسهیل بازیابی نام‌های مستعار نویسندگان ایرانی و یکدست‌سازی آن‌ها برای استفاده در «مارک ایران» (جهانشاهی ۱۳۸۵) از دیگر پژوهش‌های مهم این

1. The Virtual International Authority File (VIAF)

حوزه هستند. در تمامی این پژوهش‌ها دغدغه برگردان اسامی فارسی به زبان‌های دیگر وجود دارد و پژوهشگران تلاش کرده‌اند با ارائه راهکارهایی از این چالش مهم بکاهند. از سوی دیگر، چالش‌های برگردان نام‌ها و اسامی مختلف نه تنها میان زبان فارسی و انگلیسی، بلکه در برگردان بین زبان‌های دیگر نیز وجود دارد و پژوهشگران بسیاری به بررسی این چالش‌ها و یافتن راهکارهایی برای رفع آن‌ها پرداخته‌اند. برخی از این پژوهش‌ها با هدف افزایش بازیافت و دقت نام نویسنده در پایگاه‌های «اسکوپوس» (Kawashima, H., & Tomizawa 2015) و «پاب‌مد» (Lerchenmueller & Sorenson 2016)، ابهام‌زدایی نام نویسنده (Kim & Kim 2019) و طراحی الگوریتم‌هایی به منظور بهبود برگردان نام‌های مبتنی بر ساندکس^۱ برای زبان تایلندی (Karonboonyanan, Sornlertlamvanich 1997) و زبان پنجابی (Meknavin 1997)، (Josan & Lehal 2010; Deep & Goyal 2011) و زبان اردو (Zahid, Rao & Siddiqui 2010) انجام شده است.

همچنین، در زبان عربی که قرابت بسیاری با الفبای فارسی دارد، مسئله برگردان اسامی مورد توجه بوده است. «الاونیزان و نایت» الگوریتم برگردان بر پایه نقشه صداها و تلفظ برای برگردان نام‌های عربی به انگلیسی طراحی کرده و نشان دادند که مدل مبتنی بر املا بسیار دقیق‌تر از مدل‌های آوامحور عمل می‌کند (Al-Onaizan & Knight 2002). «یحیی» و همکاران عملکرد جست‌وجوی جدیدی بنام ساندکس عربی معرفی کردند که در جست‌وجو و بازیابی اسامی نگاشته‌شده به زبان عربی کاربرد داشت و می‌توانست در پایگاه داده کتابخانه دیجیتال ذخیره شود (Yahia et al. 2004). «فلیح» به تنوع، ناسازگاری و بی‌ثباتی برگردان کلمات انگلیسی به عربی پرداخت و دو دسته از مشکلات عمده فرایند برگردان کلمات انگلیسی را برشمرد (Falih 2009). «کی و رینر» پژوهشی درباره برگردان و تطبیق نام‌های عربی با رویکرد ترکیبی و با هدف طراحی نرم‌افزاری برای برگردان و تطبیق نام‌های عربی با دقت قابل قبول و مطابق با استاندارد نگاشتند (Kay & Rineer 2012). با مرور پژوهش‌های انجام‌شده می‌توان گفت که علی‌رغم اهمیت مسئله یکدستی برگردان اسامی در کتابخانه‌ها و پایگاه‌ها، فهرست‌های اسامی مشاهیر و مؤلفان «سازمان اسناد و کتابخانه ملی ایران» و «کتابخانه کنگره آمریکا»، و به‌خصوص پایگاه‌های استنادی

۱. Soundes: الگوریتم ساندکس، الگوریتمی است که برای نمایه کردن، به‌جای املا اصطلاحات از آوای آن‌ها استفاده می‌کند. در این روش بر روی کلمات هم‌آوا تمرکز می‌شود و علی‌رغم اختلاف جزئی در املا کلمات، آن‌ها را در یک گروه قرار می‌دهد.

مانند «وب آوساینس» و علی‌رغم تلاش‌های صورت گرفته، این مسئله همچنان به‌عنوان چالشی برای اسامی فارسی مطرح است. این مسئله در مورد سایر زبان‌ها نیز وجود دارد و تلاش‌های قابل ملاحظه‌ای در جهت رفع آن انجام گرفته است. همچنین، در مرور متون، پژوهشی که به‌طور مستقیم یا غیرمستقیم چالش‌های برگردان نام‌ها از فارسی به انگلیسی را مورد بررسی قرار داده باشد، مشاهده نشد.

۳. روش پژوهش

این پژوهش از نظر هدف، کاربردی است و از نظر شیوه گردآوری داده‌ها، توصیفی از نوع پیمایشی با رویکرد علم‌سنجی است. جامعه آماری پژوهش حاضر شامل کلیه اسامی پژوهشگران ایرانی در پایگاه «وب آوساینس» است که در بازه زمانی ۲۰۱۰ تا ۲۰۱۷ مقاله انگلیسی نگاشته‌اند. نمونه پژوهش به‌صورت هدفمند انتخاب شد. به‌دلیل این که هدف پژوهش شناسایی چالش‌های برگردان اسامی بود، نام‌هایی که دارای تنها یک صورت نگارشی بودند، با فرض این که این نام‌ها چالش‌برانگیز نیستند، حذف شده و بر این اساس تعداد ۳۹۵۹ نام منحصربه‌فرد، که دارای صورت‌های نگارشی مختلف بودند، مبنای تحلیل‌های بعدی قرار گرفت. ابزارهای به‌کاررفته در این پژوهش نرم‌افزار «یب‌اکسل»، «اکسل» و پایگاه «وب آوساینس» است. از نرم‌افزار «یب‌اکسل» برای جداسازی اسامی نویسندگان مقالات استفاده شد. همچنین، پس از استخراج اسامی توسط نرم‌افزار «یب‌اکسل»، به‌منظور پالایش و تحلیل‌های بعدی از نرم‌افزار «اکسل» استفاده شد. از پایگاه «وب آوساینس» به‌منظور استخراج اسامی و تعداد مقالات مربوط به هر اسم استفاده شد. همچنین، سیاهه واری و ارسای محقق‌ساخته متشکل از نام‌ها، صورت‌های نگارشی مختلف آن‌ها و برگردان فارسی هر یک از صورت‌ها در پژوهش به‌کار برده شد.

به‌منظور گردآوری داده‌های پژوهش ابتدا در قسمت جست‌وجوی پایگاه «وب آوساینس»، فرمول جست‌وجو بر اساس نوع مدرک (مقاله) و زبان مدرک (انگلیسی) محدود شد. سپس، نام‌های نویسندگانی که کشور آن‌ها ایران (cu=iran) بود، جست‌وجو شد. پس از آن، از اطلاعات کتابشناختی مدارک بازیابی شده در پایگاه، خروجی با فرمت «یب‌تکست»^۱ گرفته شد و فایل‌های خروجی با نرم‌افزار «یب‌اکسل» به فرمت پنجره‌ای

از نام‌ها در «اکسل» تبدیل شد. در مرحله بعد، یک نام خانوادگی واحد چندین بار در داده‌ها تکرار شده بود. در نتیجه، برای نام‌هایی که در کل فایل چندین بار تکرار شده بودند، موارد تکراری حذف شد. همچنین، تمامی اسامی که تنها یک صورت نگارشی داشتند، کنار گذاشته شد. در مرحله بعد، نام‌های با صورت‌های نگارشی مختلف در زیر هم قرار گرفت. برای این منظور، در ستونی دیگر در جلوی صورت‌های نگارشی مختلف یک نام یکی از صورت‌ها تکرار شد تا با مرتب‌سازی این ستون تمامی صورت‌های نگارشی در ستون نام‌ها در کنار یکدیگر قرار گیرند.

5516	nekui	Nekoee
5517	nekui	Nekouee
5518	nekui	Nekuee
5519	nekui	Nekoei
5520	nekui	Nekooei
5521	nekui	Nekooi
5522	nekui	Nekooie
5523	nekui	Nekouei
5524	nekui	Nekoui
5525	nekui	Nekouie

شکل ۱. نحوه قرار گرفتن نام‌ها در «اکسل»

در گام بعد، با مطالعه عمیق اسامی، نام‌ها به فارسی برگردانده شد تا تلفظ آوایی آن‌ها برای محقق روشن باشد. سپس، به هر یک از حروف، آواها و نشانه‌های دارای تلفظی واحد کد مشخص واحدی داده شد تا فارغ از صورت نگارشی مختلف آن‌ها در انگلیسی، نام‌هایی که شامل این حروف هم‌آوا بودند، در کنار یکدیگر قرار گیرند. به‌عنوان مثال، به همه صورت‌های نگارشی انگلیسی نام «نکویی» که شامل آوای (او) در فارسی بودند، کد (u) داده شد که البته، برای شناسایی صورت‌های نگارشی استفاده شده در انگلیسی، صورت نگارشی آن واج در پرانتز در جلوی آن کد قرار گرفت و به همین ترتیب، نام‌هایی که دارای آوای (او) در فارسی بودند، و برگردان انگلیسی آن‌ها دارای چند صورت نگارشی بود، کد گذاری شدند.

5516	nekui	Nekoee	i(ee),u(o)
5517	nekui	Nekouee	i(ee),u(ou)
5518	nekui	Nekuee	i(ee),u(u)
5519	nekui	Nekoei	i(ei),u(o)
5520	nekui	Nekooei	i(ei),u(oo)
5521	nekui	Nekooi	u(oo),i(i)
5522	nekui	Nekooie	i(ie),u(oo)
5523	nekui	Nekouei	i(ei),u(ou)
5524	nekui	Nekoui	u(ou),i(i)
5525	nekui	Nekouie	i(ie),u(ou)
5526	nekui	Nekuei	i(ei),u(u)
5527	nekui	Nekui	i(i),u(u)
5528	nekui	Nekuie	i(ie),u(u)

شکل ۲. کد گذاری نام‌ها در «اکسل»

به این ترتیب، اسامی دارای نگارش یکسان در فارسی که دارای صورت‌های نگارشی مختلف در انگلیسی بودند، در کنار هم قرار گرفتند و چالش‌ها و صورت‌های مختلف نگارشی آن‌ها در نام‌ها شناسایی شد و به دسته‌بندی‌های واژه‌ها، همخوان‌ها، چالش حذف و تکرار و چالش تلفظی تقسیم شد. لازم به ذکر است که برخی اسامی دارای چندین چالش بودند که به صورت جداگانه برای هر چالش کد مورد نظر آن چالش اختصاص داده شد و در نهایت، فراوانی و درصد فراوانی هر یک از چالش‌ها محاسبه شد.

۴. تجزیه و تحلیل یافته‌ها

چالش‌های برگردان نام‌های فارسی به انگلیسی و فراوانی هر یک از آن‌ها

به منظور شناسایی چالش‌های برگردان اسامی فارسی به انگلیسی شکل‌های مختلف اسامی در کنار یکدیگر قرار داده شد. سپس، واج یا آوایی را که تنوع نگارشی در آن رخ داده بود، جدا شد و صورت‌های نگارشی مختلف به عنوان مصادیق چالش آن واج در نظر گرفته شد. همچنین، کنترل شد که این صورت‌های نگارشی مربوط به تلفظ یکسانی از یک نام باشند، نه تلفظ متفاوت. در نهایت، نتایج به چهار گروه چالش‌های مربوط به «همخوان‌ها» (جدول ۱) و چالش‌های مربوط به «واژه‌ها» (جدول ۷) و چالش‌های مربوط به «حذف یا تکرار حروف» (جدول ۱۳) و چالش‌های «تلفظی» (جدول ۱۶) تقسیم گردید.

۴-۱. چالش همخوان‌ها

در برخی همخوان‌ها، نویسندگان از چند صورت نگارشی برای برگردان نام‌های فارسی به انگلیسی استفاده کرده‌اند که این همخوان‌ها (۸ مورد) مشتمل بر ۱۱ حرف است (جدول ۱). تنوع صورت‌های نگارشی از ۲ تا ۸ حالت متغیر است که در ادامه، فراوانی هر یک از چالش‌های همخوان‌ها به صورت جداگانه بیان شده است. همچنین، برای هر چالش، مصادیق آن به تفصیل آمده است.

جدول ۱. چالش‌های همخوان‌ها

ردیف	چالش	معادل حرفی در زبان فارسی	حروف معادل
۱	واج /G/	(ق، غ)	k ·c ·gh ·g ·q
۲	واج /ʒ/	(ژ)	g ·z ·j ·zh ·zh
۳	واج /ʔ/	(ع)، همزه	ea ·aa ·ei ·ee ·e ·i ·a ·ae
۴	واج /v/	(و)	w ·v
۵	واج /k/	(ک)	g ·c ·ck ·k
۶	واج /dʒ/	(ج)	g ·z ·dj
۷	واج /g/	(گ)	gh ·k ·g
۸	واج /s/	(ث، س، ص)	c ·s

۴-۱-۱. چالش واج /G/ و چالش واج /ʒ/

واج /G/ در زبان فارسی مشتمل بر دو صورت حرفی (ق، غ) است. در برگردان نام‌های فارسی که دارای واج /G/ بودند، از مصادیق q- g- gh- c- k استفاده شده است. (جدول ۲).
واج /ʒ/ در زبان فارسی مشتمل بر صورت حرفی (ژ) است که چالش واج /ʒ/ شامل نام‌هایی است که دارای واج /ʒ/ هستند و در برگردان آن‌ها به انگلیسی مصادیق j- zh- jh- g- z به کار رفته است (جدول ۲).

جدول ۲. فراوانی چالش واج /G/ و فراوانی چالش واج /J/

فراوانی چالش واج /J/			فراوانی چالش واج /G/		
درصد فراوانی	فراوانی در داده‌ها	صورت‌های تگارش‌ی	درصد فراوانی	فراوانی در داده‌ها	صورت‌های تگارش‌ی
۵۴/۰۲	۲۸۲	J	۸۵/۱۸	۱۱۶۷	gh
۴۱	۲۱۴	Zh	۷/۵۹	۱۰۴	Q
۳/۸۳	۲۰	Jh	۶/۶۴	۹۱	g
۰/۹۶	۵	G	۰/۴۴	۶	k
۰/۱۹	۱	Z	۰/۱۵	۲	c
۱۰۰/۰۰	۵۲۲	جمع	۱۰۰/۰۰	۱۳۷۰	جمع

۴-۱-۲. چالش واج /P/

واج /P/ در فارسی مشتمل بر دو صورت حرفی (ع) و همزه است که در برگردان حرف (ع) از فارسی به انگلیسی این حرف با مصوت‌ها تلفیق می‌شود. اما در این پژوهش مشاهده شد که حرف (ع) زمانی که به صورت ساکن در نام‌ها آمده، چالش برانگیز شده و برگردان این حرف به انگلیسی دارای تنوع نگارش‌ی ae- ea- e- i- ei- ee- aa- a- شده است (جدول ۳).

جدول ۳. فراوانی چالش حرف (ع)

صورت‌های تگارش‌ی	فراوانی در داده‌ها	درصد فراوانی
A	۷۶	۳۳/۹۳
Aa	۵۶	۲۵
E	۲۶	۱۱/۶۱
Ei	۲۳	۱۰/۲۷
Ee	۲۱	۹/۳۷
I	۱۱	۴/۹۱
Ae	۶	۲/۶۸
Ea	۵	۲/۲۳
جمع	۲۲۴	۱۰۰/۰۰

۴-۱-۳. چالش واج /v/

واج /v/ در زبان فارسی مشتمل بر صورت حرفی (و) است. بررسی نام نویسندگان در پایگاه نشان داد که در برگردان واج /v/ به انگلیسی صورت‌های w-v نگاشته شده است. نتایج نشان داد که در چالش واج /v/ صورت نگارشی v با ۵۱/۹۶ درصد فراوانی، از فراوانی بیشتری نسبت به صورت نگارشی w برخوردار است. اما به این دلیل که در این پژوهش فقط نام‌هایی که دارای چالش بودند بررسی شده، فراوانی دو صورت نگارشی به هم نزدیک شده‌اند (جدول ۴).

جدول ۴. فراوانی چالش واج /v/

درصد فراوانی	فراوانی در داده‌ها	صورت‌های نگارشی
۵۱/۹۶	۵۳	v
۴۸/۰۴	۴۹	w
۱۰۰/۰۰	۱۰۲	جمع کل

۴-۱-۴. چالش واج /k/ و چالش واج /dʒ/

واج /k/ در زبان فارسی مشتمل بر صورت حرفی (ک) است که برخی نویسندگان و نگارندگان واج /k/ را در برگردان به انگلیسی به صورت‌های g-ck-c نیز می‌نگارند. در چالش واج /k/ صورت نگارشی k بیشترین فراوانی را در مقایسه با دیگر صورت‌های نگارشی این واج دارد (جدول ۵).

واج /dʒ/ در زبان فارسی مشتمل بر صورت حرفی (ج) است که بررسی نام‌ها در پایگاه «وب‌آوساینس» نشان داد که در برگردان نام‌هایی که دارای واج /dʒ/ بودند، صورت‌های dj-g از استفاده شده بود. با توجه به فراوانی داده‌ها در برگردان واج /dʒ/ می‌توان گفت که صورت‌های نگارشی z و dz بیشتر از دیگر صورت‌ها مورد استفاده قرار گرفته است (جدول ۵).

جدول ۵. فراوانی چالش واج /k/ و فراوانی چالش واج /dʒ/

فراوانی چالش واج /dʒ/			فراوانی چالش واج /k/		
درصد فراوانی	فراوانی در داده‌ها	صورت‌های نگارشی	درصد فراوانی	فراوانی در داده‌ها	صورت‌های نگارشی
۴۵/۸۳	۲۲	J	۵۱/۱۱	۴۶	K
۳۵/۴۲	۱۷	Dz	۳۰	۲۷	Ck
۱۸/۷۵	۹	G	۱۳/۳۳	۱۲	G
۱۰۰/۰۰	۴۸	جمع	۵/۵۶	۵	C
			۱۰۰/۰۰	۹۰	جمع

۴-۱-۵. چالش واج /g/ و چالش واج /s/

واج /g/ در زبان فارسی مشتمل بر صورت حرفی (گک) است که در برگردان این واج، برخی نویسندگان از صورت نگارشی k استفاده کرده‌اند (جدول ۶).
واج /s/ در زبان فارسی مشتمل بر سه صورت حرفی (س، ص، ث) است. در برگردان نام‌هایی که دارای حروف (س، ص، ث) هستند، انتظار می‌رود که صورت‌های نگارشی s، c، th برای برگردان آن‌ها استفاده شود، اما در یافته‌های این پژوهش مشخص شد که در نمونه‌ای که از نام‌های نویسندگان پایگاه گرفته شده از صورت‌های s-c استفاده شده است (جدول ۶).

جدول ۶. جدول فراوانی چالش واج /g/ و فراوانی چالش واج /s/

فراوانی چالش واج /s/			جدول فراوانی چالش واج /g/		
درصد فراوانی	فراوانی در داده‌ها	صورت‌های نگارشی	درصد فراوانی	فراوانی در داده‌ها	صورت‌های نگارشی
۶۴/۲۹	۹	S	۶۵/۲۲	۳۰	G
۳۵/۷۱	۵	C	۳۲/۶۱	۱۵	k
۱۰۰/۰۰	۱۴	جمع	۲/۱۷	۱	gh
			۱۰۰/۰۰	۴۶	جمع

۴-۲. چالش واکه‌ها

خط فارسی از نوع ابجد است و با الفبا تفاوت دارد. در خط ابجدی، تنها برای همخوان‌ها (بی‌آواها) واج وجود دارد و واکه‌ها (آوادارها) یا نوشته نمی‌شوند یا با

جدول ۸. فراوانی چالش واکه /i/

ردیف	صورت‌های نگارشی	فراوانی در داده‌ها	درصد فراوانی	ردیف	صورت‌های نگارشی	فراوانی در داده‌ها	درصد فراوانی
۱	i	۱۷۰۳	۲۹/۹۹	۱۰	Yi	۶۷	۱/۱۸
۲	ei	۱۱۵۱	۲۰/۲۷	۱۱	lee	۴۸	۰/۸۴
۳	y	۹۶۸	۱۷/۰۴	۱۲	lei	۳۱	۰/۵۵
۴	ee	۵۱۳	۹/۰۳	۱۳	Yee	۲۰	۰/۳۵
۵	ie	۳۸۱	۶/۷۱	۱۴	Eei	۱۴	۰/۲۵
۶	ey	۳۲۹	۵/۷۹	۱۵	Eie	۱۰	۰/۱۸
۷	iy	۲۱۰	۳/۷۰	۱۶	Ye	۹	۰/۱۶
۸	e	۱۱۳	۱/۹۹	۱۷	ley	۵	۰/۰۹
۹	ii	۱۰۷	۱/۸۸	۱۸	جمع	۵۶۷۹	۱۰۰/۰۰

۴-۲-۲. چالش واکه /u/

واج /u/ در زبان فارسی با صورت حرفی (او) نمایش داده می‌شود. واکه /u/ در برگردان نام‌های فارسی به انگلیسی دارای صورت‌های نگارشی o- u- ou- oo- است. نتایج به‌دست آمده از بررسی فراوانی چالش واکه /u/ نشان داد که صورت نگارشی ou و oo بیشترین فراوانی نگارشی را دارد (جدول ۹).

جدول ۹. فراوانی چالش واکه /u/

صورت‌های نگارشی	فراوانی در داده‌ها	درصد فراوانی
Ou	۱۸۴۵	۴۰/۴۵
oo	۱۵۳۱	۳۳/۵۷
o	۵۹۲	۱۲/۹۸
u	۵۶۵	۱۲/۳۹
uo	۲۸	۰/۶۱
جمع	۴۵۶۱	۱۰۰/۰۰

۴-۲-۳. چالش توالی واج‌های /e/ و /z/

این چالش شامل واژه‌هایی می‌شود که دارای توالی دو واج /e/ و /z/ هستند که در برگردان آن‌ها از فارسی به انگلیسی صورت‌های ai- ei- ey- e- i- ay- y- a- ie- iy- ea مورد

استفاده قرار گرفته است. نویسندگان مقالات موجود در پایگاه، اغلب نام‌های واجد این توالی را به صورت (ei) می‌نگارند (جدول ۱۰).

جدول ۱۰. فراوانی چالش توالی واج‌های /e/ و /a/

ردیف	صورت‌های نگارشی	فراوانی در داده‌ها	درصد فراوانی
۱	Ei	۴۲۱	۴۲/۴۸
۲	ey	۲۸۵	۲۸/۷۶
۳	ai	۹۲	۹/۲۸
۴	ay	۸۳	۸/۳۷
۵	i	۴۹	۴/۹۴
۶	y	۲۶	۲/۶۲
۷	E	۱۵	۱/۵۱
۸	le	۱۱	۱/۱۱
۹	A	۵	۰/۵۰
۱۰	ly	۴	۰/۴۰
	جمع	۹۹۱	۱۰۰/۰۰

۴-۲-۴. چالش پایانه (ه)

یکی از صورت‌های حرفی واکه /e/ صورت (ه) است که به آن «های» غیرملفوظ گفته می‌شود. به این دلیل که /h/ تلفظ نمی‌شود و /e/ تلفظ می‌شود. نویسندگان نگارندگان برای برگردان واکه (ه) در آخر نام صورت‌های e-eh را به کار برده‌اند.

جدول ۱۱. فراوانی چالش پایانه (ه)

صورت‌های نگارشی	فراوانی در داده‌ها	درصد فراوانی
eh	۴۹۵	۶۳/۷۱
e	۲۸۲	۳۶/۲۹
جمع	۷۷۷	۱۰۰/۰۰

۴-۲-۵. چالش واکه /o/

واکه /o/ در زبان فارسی معادل حرفی ندارد. در برگردان آن دسته از نام‌های

نویسندگان دارای واژه /o/ در پایگاه، صورت‌های نگارشی oo- ou- u- o- استفاده شده است (جدول ۱۲).

جدول ۱۲. فراوانی چالش واژه /o/

صورت‌های نگارشی	فراوانی در داده‌ها	درصد فراوانی
o	۱۱۸	۳۸/۰۶
ou	۷۴	۲۳/۸۷
ow	۴۹	۱۵/۸۱
u	۴۷	۱۵/۱۶
oo	۲۰	۶/۴۵
ov	۲	۰/۶۵
جمع	۳۱۰	۱۰۰/۰۰

۳-۴. چالش‌های حذف یا تکرار حروف

به دلیل این که این چالش‌ها در دو گروه چالش‌های همخوان‌ها و واژه‌ها نمی‌گنجیدند، به صورت جداگانه مورد بحث قرار گرفت (جدول ۱۳). چالش حذف یا تکرار حروف شامل تشدید و حذف ال می‌شود. در این بخش فراوانی مصادیق هر یک از چالش‌های تشدید، حذف (ال) بررسی شده است.

جدول ۱۳. چالش‌های تشدید، حذف، ال

ردیف	چالش	حروف معادل
۱	تشدید	zz, yy, vv, tt, ss, rr, pp, nn, mm, ll, kk, jj, hh, ff, dd, bb
۲	(ال)	ال دارد- ال ندارد

۳-۴-۱. چالش تشدید

چالش تشدید^۱ چالشی است که در آن نویسندگان و نگارندگان در برگردان نشانه تشدید به انگلیسی گاهی واجی را که در فارسی دارای تشدید است، تکرار می‌کنند و

1. double consonant

گاهی آن را تکرار نمی‌کنند. در ۱۶ حرف از ۲۶ حرف برگردان الفبای فارسی به انگلیسی تشدید اتفاق افتاده است. تشدید در حروف a, c, e, g, o, q, u, w صورت نگرفته است. پس از بررسی حروفی که در آن‌ها تشدید اتفاق افتاده بود، این نتیجه حاصل شد که صورت مشدد حروف در نام‌های نویسندگان در کل، نسبت به صورت غیرمشدد همان نام‌ها فراوانی بیشتری دارند. در عین حال، در برخی حروف مانند h, k, n, t, v, y, z بین فراوانی صورت مشدد و صورت غیرمشدد این حروف تفاوت چشمگیری وجود ندارد (جدول ۱۴).

جدول ۱۴. فراوانی چالش تشدید

صورت‌های نگارشی	بدون تشدید		با تشدید	
	درصد فراوانی	فراوانی در داده‌ها	درصد فراوانی	فراوانی در داده‌ها
s	۲۶/۲۲	۲۵۸	۲۳/۳۸	۲۹۲
l	۱۳/۰۱	۱۲۸	۱۸/۳۳	۲۲۹
d	۱۰/۷۷	۱۰۶	۱۰/۹۷	۱۳۷
m	۷/۹۳	۷۸	۸/۵۷	۱۰۷
b	۷/۱۱	۷۰	۷/۹۳	۹۹
r	۸/۰۳	۷۹	۷/۶۹	۹۶
y	۶/۶۱	۶۵	۵/۲۸	۶۶
f	۴/۴۷	۴۴	۴/۴۸	۵۶
t	۳/۶۶	۳۶	۲/۹۶	۳۷
z	۲/۹۵	۲۹	۲/۴۰	۳۰
j	۲/۵۴	۲۵	۲/۲۴	۲۸
v	۲/۲۴	۲۲	۱/۸۴	۲۳
h	۱/۷۳	۱۷	۱/۶۰	۲۰
n	۱/۶۳	۱۶	۱/۲۸	۱۶
k	۱/۱۲	۱۱	۰/۸۸	۱۱
p	۰	۰	۰/۱۶	۲
جمع	۱۰۰/۰۰	۹۸۴	۱۰۰/۰۰	۱۲۴۹

۴-۳-۲. چالش حذف (ال)

برخی از نام‌ها دارای (ال) هستند که برخی نویسندگان و نگارندگان در برگردان این نام‌ها به انگلیسی این (ال) را حذف می‌کنند؛ مانند نام عبدالرسولی Abdolrasouli که به صورت Abdorasouli، و همچنین، نام صدرالدینی Sadroddini که به صورت Sadroddini نگاشته شده است.

جدول ۱۵. فراوانی چالش حذف (ال)

مصادیق	فراوانی در داده‌ها
دارای (ال)	۸۷
بدون (ال)	۲۵
جمع	۱۱۲

۴-۴. چالش‌های تلفظی

این دسته شامل واژه /æ/، /e/ و حذف برخی حروف می‌شود که به دلیل تفاوت در تلفظ این حروف در بین نویسندگان مختلف موجب به وجود آمدن نام‌های چند نگارشی شده است (جدول ۱۶).

جدول ۱۶. چالش‌های تلفظی

ردیف	چالش	حروف معادل
۱	واژه /æ/	i, u, o, e, a
۲	واژه /e/	a, i, u, o, e
۳	حذف	a, o, y, e, h

۴-۴-۱. چالش واژه /æ/ و چالش واژه /e/

واژه /æ/ در زبان فارسی صورت حرفی ندارد. در برگردان از فارسی به انگلیسی نام‌هایی که شامل واژه /æ/ هستند، این واژه به صورت‌های i - u - o - e - a به کار رفته است. مانند نام‌های «لنگرودی» که به صورت‌های Langeroodi، Langaroodi، «معارف» به صورت‌های Moaref، Maaref، «عبداللهی» به صورت‌های Abdollahi، Abdullahi، Abdollahi و «اسماعیلی» که به صورت‌های Esmaili، Ismaili نگاشته شده است. از بین صورت‌های

نگارشی برگردان واکه /æ/ از فارسی به انگلیسی، صورت نگارشی a بیشترین فراوانی را دارد (جدول ۱۷).

واج /e/ در زبان فارسی معادل حرفی ندارد. این واکه به صورت‌های a-e-o-i-u در برگردان نام‌ها از فارسی به انگلیسی مورد استفاده قرار گرفته است؛ مانند نام‌های «اردبیلی» که به صورت‌های Ardebili، Ardabili، «نجم‌الدین» که به صورت‌های Najmodin، Najmedin و نام «یونس» که به صورت‌های Yunis، Yunes، Yunus نگاشته شده‌اند. نتایج حاصل از بررسی فراوانی چالش واکه /e/ نشان داد که صورت‌های نگارشی e و a در قیاس با دیگر صورت‌های نگارشی فراوانی بیشتری را به خود اختصاص داده‌اند (جدول ۱۷).

جدول ۱۷. فراوانی چالش واکه /æ/ و فراوانی چالش واکه /e/

فراوانی چالش واکه /e/			فراوانی چالش واکه /æ/		
درصد فراوانی	فراوانی در داده‌ها	صورت‌های نگارشی	درصد فراوانی	فراوانی در داده‌ها	صورت‌های نگارشی
۴۲/۲۴	۱۴۷	e	۳۶/۳۶	۱۵۲	a
۳۳/۹۱	۱۱۸	a	۳۴/۲۱	۱۴۳	e
۱۸/۳۹	۶۴	o	۲۴/۴۰	۱۰۲	o
۲/۸۷	۱۰	i	۳/۵۹	۱۵	u
۲/۵۹	۹	u	۱/۴۴	۶	i
۱۰۰/۰۰	۳۴۸	جمع	۱۰۰/۰۰	۴۱۸	جمع

۴-۲. چالش حذف

چالش حذف چالشی است که در آن به دلیل وجود تلفظ‌های مختلف برای یک نام، برخی حروف از برخی نام‌ها ممکن است حذف شود. به عنوان مثال، در حذف e نام «لنگرودی» langeroodi به صورت langroodi و نام «لاجوردی» Lajevardi با صورت نگارشی Lajvardi نیز برگردان می‌شود. در حذف y نام «لیاقت» به شکل‌های نگارشی Liaghat، Liyaghat همچنین، نام «شایقی» به صورت‌های Shayeghi، Shaeghi نگاشته می‌شود. در حذف a نام «ترکمان» به صورت‌های نگارشی Torkman، Torkaman، و نام «مهربان» که به شکل‌های نگارشی Mehrban، Mehrban برگردان می‌شود. در حذف o نام «بلوکی» به صورت‌های Boluki، Bluki، و نام «ارجمندی» با مصادیق Arjomandi، Arjmandi نگاشته شده است. در نهایت، در حذف واج h نام «مشهدی‌زاده» با مصادیق Mashadizadeh،

Mashhadizadeh و نام «منوچهری» به صورت‌های Manoochehri، Manoucheri برگردان شده است.

جدول ۱۸. فراوانی چالش حذف

مصادیق	با حذف		بدون حذف	
	فراوانی در داده‌ها	درصد فراوانی	فراوانی در داده‌ها	درصد فراوانی
e	۷۹	۴۳/۴۱	۹۱	۳۷/۱۴
a	۲۴	۱۳/۱۹	۳۹	۱۵/۹۲
h	۳۱	۱۷/۰۳	۳۳	۱۳/۴۷
o	۲۷	۱۴/۸۴	۵۲	۲۱/۲۲
y	۲۱	۱۱/۵۴	۳۰	۱۲/۲۴
جمع	۱۸۲	۱۰۰/۰۰	۲۴۵	۱۰۰/۰۰

۵. بحث

پس از بررسی صورت‌های نگارشی نام‌های نویسندگان و فراوانی هر یک از صورت‌های نگارشی در داده‌ها، چالش‌های برگردان نام‌ها در چهار دسته چالش‌های همخوان‌ها، چالش‌های واکه‌ها و چالش‌های حذف و تکرار، و چالش‌های تلفظی تقسیم‌بندی شدند که در ادامه، ابتدا نتایج بررسی هر چالش و صورت‌های نگارشی آن به‌طور خلاصه جمع‌بندی خواهد شد و پس از آن با جداول آوانگاری بین‌المللی، جدول دادگان، و جدول دگرنویسی «کتابخانه کنگره»^۱ به‌عنوان سه مرجع مقایسه خواهند شد. در دسته اول، چالش‌های همخوان‌ها، در نتایج بررسی برگردان واج /G/ مشخص شد که، صورت نگارشی gh رفتار نگارشی غالب نویسندگان و نگارندگان است (جدول ۲). به‌منظور مقایسه بهتر جدول آوانگاری بین‌المللی^۲ برای برگردان واج /G/ صورت نگارشی

۱. جدول آوانگاری بین‌المللی، رابطه الفبای بین‌المللی آوانگاری و آواهای زبان فارسی را بررسی کرده، اما جدول دادگان رابطه واج‌های فارسی را با الفبای زبان انگلیسی در محیط‌های رایانه‌ای نشان داده است. جدول دگرنویسی «کتابخانه کنگره»، الگوی دگرنویسی مورد استفاده در «کتابخانه کنگره» را ارائه کرده است.

2. IPA (International Phonetic Alphabet)

g و جدول دادگان فارسی^۱ برای برگردان این واج صورت نگارشی q را پیشنهاد داده‌اند. علاوه بر این جدول دگرنویسی «کتابخانه کنگره» برای برگردان این واج صورت‌های نگارشی q، gh، g در نظر گرفته است. همان‌گونه که مشاهده می‌شود، شکل غالب نگارشی نویسندگان در مورد واج /G/ با جدول بین‌المللی آوانویسی و حتی دادگان فارسی همخوانی ندارد.

در برگردان واج /ʒ/ از فارسی به انگلیسی، صورت‌های نگارشی z و zh شکل نگارشی غالب در برگردان فارسی به انگلیسی واج /ʒ/ در بین نویسندگان و نگارندگان است (جدول ۲). در جدول آوانگاری بین‌المللی صورت نگارشی برگردان واج /ʒ/، با نشانه "ʒ"، در جدول دادگان فارسی صورت نگارشی Z و جدول دگرنویسی «کتابخانه کنگره» صورت نگارشی zh پیشنهاد شده است. همان‌گونه که مشاهده می‌شود، شکل غالب نگارشی نویسندگان در مورد واج /ʒ/ با جدول بین‌المللی آوانویسی و حتی دادگان فارسی همخوانی ندارد.

همچنین، در برگردان حرف (ع)، مشخص شد که صورت نگارشی e، صورت نگارشی غالب در برگردان واج /ʔ/ از فارسی به انگلیسی است (جدول ۳). این نتیجه پژوهش با پیشنهاد جدول دگرنویسی «کتابخانه کنگره» همخوانی دارد. با وجود این، جدول آوانگاری بین‌المللی برای برگردان این واج نشانه "ʔ" و جدول دادگان فارسی نشانه «ʔ» را ارائه کرده‌اند. در نتایج بررسی برگردان واج /v/ مشخص شد که صورت نگارشی v، رفتار نگارشی غالب نویسندگان و نگارندگان است (جدول ۴). دلیل این امر را می‌توان در نبود واژه w در فارسی، برخلاف زبان‌های عربی و انگلیسی دانست. به‌منظور مقایسه بهتر جدول آوانگاری بین‌المللی برای برگردان واج /v/ صورت نگارشی v و جدول دادگان فارسی نیز برای برگردان این واج صورت نگارشی v را پیشنهاد داده‌اند. اما جدول دگرنویسی «کتابخانه کنگره» برای برگردان این واج صورت‌های نگارشی v و w را در نظر گرفته است. در برگردان واج /k/ از فارسی به انگلیسی در نام‌ها از مقایسه صورت‌های نگارشی و فراوانی هر یک از آن‌ها مشخص شد که صورت نگارشی k رفتار نگارشی غالب نویسندگان و نگارندگان است (جدول ۵). در جدول آوانگاری بین‌المللی، جدول دادگان

۱. جدول نشانه‌های ساده‌شده آوانویسی فارسی: در تدوین این جدول تلاش شده تا همه واج‌های فارسی با نشانه‌های ساده الفبای انگلیسی بدون هیچ تغییر شکل یا نشانه اضافی نمایش داده شوند تا در همه محیط‌های رایانه‌ای یکسان دیده شوند.

فارسی و جدول دگرنویسی «کتابخانه کنگره» نیز صورت نگارشی k برای برگردان واج /k/ از فارسی به انگلیسی پیشنهاد شده است.

از مقایسه صورت‌های نگارشی مختلف در برگردان واج /dʒ/ از فارسی به انگلیسی مشخص شد که صورت‌های نگارشی زو و dj، صورت نگارشی غالب نویسندگان و نگارندگان هستند (جدول ۵). همچنین، جدول دگرنویسی «کتابخانه کنگره» نیز صورت‌های نگارشی زو و dj را در برگردان واج /dʒ/ در نظر گرفته است که با پژوهش «صدیق‌بهزادی» (۱۳۷۱) همسوست. به‌منظور مقایسه بهتر جدول آوانگاری بین‌المللی برای برگردان واج /dʒ/ نشانه dʒ و جدول دادگان فارسی برای برگردان این واج صورت نگارشی را پیشنهاد داده‌اند. همچنین، مقایسه نتایج پژوهش حاضر و جدول آوانگاری بین‌المللی همخوانی ندارند. البته، این احتمال وجود دارد که علت استفاده از صورت نگارشی dj آن است که، d و j در کنار هم صدای /dʒ/ را می‌دهد، ولی ز به تنهایی صدای /ʒ/ می‌دهد و برای اجتناب از آوای دوم، معمولاً d را با ز به کار می‌برند؛ کما این که این سنت در زبان فرانسه نیز دیده می‌شود. از مقایسه صورت‌های نگارشی در برگردان واج /g/ از فارسی به انگلیسی مشخص شد که صورت نگارشی g رفتار نگارشی غالب نویسندگان و نگارندگان است (جدول ۶). در جدول آوانگاری بین‌المللی، جدول دادگان فارسی و جدول دگرنویسی «کتابخانه کنگره» نیز صورت نگارشی g برای برگردان واج /g/ از فارسی به انگلیسی پیشنهاد شده است که نتایج پژوهش با داده‌های این جداول همخوانی دارد.

همان‌گونه که «گل‌تاجی و بذرگر» بیان کرده‌اند، یکی از ویژگی‌های زبان فارسی این است که برای یک آوا چند علامت مختلف داریم؛ مانند علامت‌های (ث، س، ص) که هر سه در فارسی یکسان خوانده می‌شوند (۱۳۸۹). از مقایسه صورت‌های نگارشی مشخص شد که صورت نگارشی s، رفتار نگارشی غالب نویسندگان و نگارندگان است (جدول ۶). در جدول آوانگاری بین‌المللی، جدول دادگان فارسی و جدول دگرنویسی «کتابخانه کنگره» نیز صورت نگارشی s برای برگردان این حروف از فارسی به انگلیسی پیشنهاد شده است که نتایج پژوهش با داده‌های این جداول همخوانی دارد.

با توجه به موارد گفته‌شده در برگردان واج بی‌آوا از فارسی به انگلیسی می‌توان نتیجه گرفت که در برگردان این حروف صورت‌های نگارشی غالب نویسندگان مطابقت چندانی با نشانه‌های پیشنهادشده در دستورالعمل‌ها ندارند که احتمال می‌رود این تضاد به علت پیروی رفتار نویسندگان از اصل کمترین کوشش و ساده‌نویسی است. به عبارت

دیگر، نویسندگان چندان مقید به قواعد توصیه‌شده نیستند و به نظر می‌رسد که عادت‌ها و عرف نقش مهم‌تری ایفا می‌کنند.

در دسته دوم (چالش واکه‌ها)، صورت‌های نگارشی مختلفی در برگردان واکه‌های /i/، /u/، توالی واج‌های /e/ و /a/، /o/ و پایانه (به) از فارسی به انگلیسی مشاهده شد که وجود این صورت‌های نگارشی مختلف در نام‌ها منجر به عدم یکدستی در بین نویسندگان و نگارندگان شده است.

از مقایسه فراوانی صورت‌های نگارشی واکه /i/ در داده‌های پژوهش مشخص شد که صورت نگارشی آ‌بیشترین فراوانی را در بین دیگر صورت‌های نگارشی به خود اختصاص داده است (جدول ۸). جدول آوانگاری بین‌المللی و جدول دادگان فارسی نیز در برگردان واکه /i/ صورت نگارشی آ‌را استفاده کرده‌اند. پس در نتیجه، یافته‌های پژوهش با جدول دادگان و جدول آوانگاری بین‌المللی همخوانی دارد. اما جدول دگرنویسی «کتابخانه کنگره» در برگردان واکه /i/ صورت‌های نگارشی ey، i، iy را در نظر گرفته است که در این پژوهش صورت نگارشی ey تنها ۵ درصد و صورت نگارشی iy سه درصد از فراوانی را دارا هستند. در برگردان واکه /u/ از فارسی به انگلیسی، پس از مقایسه فراوانی صورت‌های نگارشی مشخص شد که صورت نگارشی ou در برگردان واکه /u/ از فارسی به انگلیسی، صورت نگارشی غالب در بین نویسندگان و نگارندگان است (جدول ۹). در جدول آوانگاری بین‌المللی و جدول دادگان فارسی برای برگردان واکه /u/ از فارسی به انگلیسی صورت‌های نگارشی u ثبت شده است. در نتیجه، یافته‌های پژوهش با جدول آوانگاری بین‌المللی و جدول دادگان فارسی همخوانی ندارد. همچنین، در جدول دگرنویسی «کتابخانه کنگره» صورت‌های نگارشی o، ow و ou برای برگردان واکه /u/ در نظر گرفته شده است. البته، این احتمال وجود دارد که جدول «کنگره» این نشانه‌ها را برای واج توصیه کرده است نه برای آوا.

نتایج پژوهش نشان داد که در برگردان توالی واج‌های /e/ و /j/ از فارسی به انگلیسی صورت نگارشی ei، صورت نگارشی غالب در بین نویسندگان و نگارندگان است (جدول ۱۰).

پایانه (به) شامل نام‌هایی است که در انتهای آن‌ها واکه /e/ و حرف (ه) وجود دارد. نتایج این پژوهش نشان داد نام‌هایی که در آن‌ها حرف (ه) حذف نشده و به عبارتی، دارای صورت نگارشی eh هستند، به‌عنوان صورت نگارشی غالب تلقی می‌شوند (جدول ۱۱). در

جدول آوانگاری بین‌المللی برای برگردان پایانه (به صورت نگارشی e در نظر گرفته شده است که با نتایج پژوهش مغایر است.

از مقایسه فراوانی صورت‌های نگارشی واکه /o/ این نتیجه حاصل شد که صورت نگارشی o در برگردان واکه /o/ از فارسی به انگلیسی، صورت نگارشی غالب در برگردان این واکه است (جدول ۱۲). جدول آوانگاری بین‌المللی و جدول دادگان فارسی نیز در برگردان واکه /o/ از فارسی به انگلیسی صورت نگارشی o را ثبت کرده‌اند که با داده‌های پژوهش حاضر همخوانی دارد. اما جدول دگرنویسی «کتابخانه کنگره» علاوه بر صورت نگارشی o صورت نگارشی u را نیز انتخاب کرده است. البته، باز هم این احتمال وجود دارد که جدول دگرنویسی «کنگره»، هر دو واکه ناظر بر یک واج را لحاظ کرده و محتمل می‌داند. به دلیل این که جدول «کنگره» بر آوا کار کرده است نه بر کلمه، آواهای ممکن برای یک واج را بیان کرده است.

در نهایت، می‌توان نتیجه گرفت که یافته‌های پژوهش در این بخش (واکه‌ها) با بسیاری از پژوهش‌هایی که از منظر این چالش‌ها بررسی شده‌اند، همسویی دارد. همچنین، می‌توان مشاهده کرد که رفتار نویسندگان و نگارندگان در واکه‌ها نسبت به حروف بی‌آوا قاعده‌مندی بیشتری داشته و با قوانین تجویز شده مطابقت بیشتری دارد.

در دسته سوم، وجود حروف مشدد و حروف (ال) در برخی نام‌ها موجب به وجود آمدن چندگانگی در نگارش نام‌ها شده است. پس از مقایسه فراوانی صورت مشدد حروف در نام‌های نویسندگان و صورت غیرمشدد همان نام‌ها مشخص شد که در کل، صورت نگارشی مشدد حروف فراوانی بیشتری نسبت به حروف غیرمشدد دارد (جدول ۱۴). همچنین، از مقایسه نتایج پژوهش و جدول آوانگاری بین‌المللی این نتیجه به دست آمد که نتایج پژوهش با جدول آوانگاری بین‌المللی همخوانی ندارد؛ در جدول آوانگاری بین‌المللی برای برگردان تشدید نشانه کشیدگی /:/ ثبت شده است، اما در پژوهش حاضر مشخص شد که تشدید در برگردان انگلیسی به صورت تکرار واجی که در فارسی دارای تشدید است به کار رفته است. علت این ناهمخوانی می‌تواند نامأنوس بودن نماد استفاده‌شده و ناآشنایی نویسندگان و نگارندگان با این قواعد باشد.

همچنین، در برگردان برخی نام‌ها به انگلیسی ممکن است املاهای متفاوتی برای این نام‌ها به کار برده شود که گاهی با (ال) و گاهی بدون (ال) نگارش شوند. نتایج پژوهش نشان داد که صورت نگارشی نام‌ها با (ال) از فراوانی بیشتری نسبت به صورت نگارشی

بدون (ال) برخوردار است. در نتیجه، شکل نگارشی نام‌ها با (ال) شکل نگارشی غالب است (جدول ۱۵).

در این دسته آواهای *ا* و *اِ* و همچنین حذف برخی حروف به علت تفاوت در تلفظ نام‌ها در گویش‌ها و زبان‌های مختلف دچار چندگانگی شده‌اند (جدول ۱۶).

نتایج پژوهش حاضر نشان داد که صورت نگارشی *a* در برگردان واژه */æ/* از فارسی به انگلیسی، صورت نگارشی غالب است (جدول ۱۷). پژوهش‌های «رائی ساربانقلی» (۱۳۸۴) و «یارمحمدی» (۱۳۶۴) نیز صورت نگارشی *a* را در برگردان واژه */æ/* غالب دانسته‌اند و با پژوهش حاضر همسو هستند. جدول دادگان فارسی و جدول دگرنویسی «کتابخانه کنگره» نیز برای برگردان واژه */æ/* صورت نگارشی *a* را ثبت کرده‌اند. اما جدول آوانگاری بین‌المللی برای برگردان واژه */æ/* نشانه *æ* را پیشنهاد می‌دهد.

در برگردان واژه */e/* از فارسی به انگلیسی از مقایسه فراوانی صورت‌های نگارشی مشخص شد که صورت نگارشی *e* در برگردان واژه */e/*، صورت نگارشی غالب در بین نویسندگان و نگارندگان است (جدول ۱۷). نتایج پژوهش حاضر با نتایج پژوهش‌های «رائی ساربانقلی» (۱۳۸۴) و «یارمحمدی» (۱۳۶۴) همسوست، زیرا در انتخاب صورت نگارشی *e* برای برگردان واژه */e/* از فارسی به انگلیسی اتفاق نظر وجود دارد. همچنین، جدول آوانگاری بین‌المللی و جدول دادگان فارسی نیز در برگردان واژه */e/* از فارسی به انگلیسی صورت نگارشی *e* را ثبت کرده‌اند و با نتایج پژوهش حاضر همخوانی دارد. اما جدول دگرنویسی «کتابخانه کنگره» صورت نگارشی *ā* را برای برگردان این واژه برگزیده است. پس از بررسی چالش‌های واژه */e/* و */æ/* یکی از دلایل بسیار مهم نامنطبق بودن صورت‌های نگارشی این واژه‌ها با صورت‌های نگارشی پیشنهادشده در استانداردها، نبود شناخت کافی در بین نویسندگان نسبت به واژه‌ای است که هر واج تولید می‌کند. به‌عنوان مثال، اگر نویسندگان و نگارندگان می‌دانستند که واج */i/* واژه (ئی) کوتاه نزدیک به (-) را تولید می‌کند مسلماً از واج */ee/* که واژه (ئی) بلند را تولید می‌کند، استفاده می‌کردند. بنابراین، ناآشنایی با تلفظ صحیح حروف و آواهای ناظر بر آن‌ها در زبان مقصد بسیار مهم است. این مسئله را می‌توان در واژه */æ/* مشاهده کرد. و به همین دلیل است که در برگردان برخی حروفی که نویسنده با واژه آن آشنایی دارد یا چندواژه‌ای نیستند، یکدستی و انطباق با قواعد پیشنهادی بیشتر مشاهده می‌شود.

بخش دیگری از یافته‌های پژوهش نشان داد که در برگردان نام‌ها از فارسی به

انگلیسی برخی حروف حذف می‌شوند که بیشتر به دلیل تفاوت تلفظ این نام‌ها در فرهنگ‌ها و گویش‌های متفاوت در کشور است. در مجموع، می‌توان نتیجه گرفت که حذف در این نام‌ها با قاعده‌های زیر اتفاق افتاده است: ۱) نشانه‌های *-و* به همراه یک صامت در برخی نام‌ها، ۲) حرف *ئی* + الف که تولید «یا» می‌کند، و ۳) قرار گرفتن حرف «ه» بعد از *ش/چ* که باعث می‌شود دو واج */h/* در انگلیسی کنار هم قرار بگیرد. همچنین، نتایج پژوهش نشان داد که به‌طور کلی، شکل نگارشی بدون حذف حروف، صورت نگارشی غالب در بین نویسندگان و نگارندگان بوده است (جدول ۱۸).

به‌طور کلی، می‌توان بیان کرد که گرچه پژوهشی که به بررسی چالش‌های برگردان در نام‌ها و در سایر موارد مشابه پرداخته باشد مشاهده نشد، اما در معدودی از پژوهش‌ها چالش‌های زبان فارسی به‌طور کلی مورد بررسی قرار گرفته است. ریشه بروز این چالش‌ها را می‌توان در مراحل مختلف چرخه حیات یک مدرک علمی از مرحله تایپ هنگام تولید مدرک تا آخرین مرحله که دروندی عبارت جست‌وجوست، یافت. همچنین، نبود استاندارد نگارش فارسی و در نتیجه، سلیقه‌ای رفتار کردن نویسندگان، نبود صفحه‌کلید و کدهای استاندارد، عادت به آسان‌نویسی و رعایت نکردن پیچیدگی‌های نگارشی هنگام تایپ عوامل به‌وجود آمدن چند دستگی در نگارش نام‌ها هستند. همچنین، اشتباهات املائی که توسط نویسندگان در نگارش نام‌ها رخ می‌دهد، منجر به ایجاد صورت‌های نگارشی مختلف برای یک نام واحد می‌شود. «عبداللهی نورعلی» (۱۳۸۶) و «محقق‌زاده و زارعیان» (۱۳۸۳) در بررسی چالش‌های زبان فارسی به‌طور کلی، و همچنین «گل‌تاجی و عباس‌پور» (۱۳۹۶) در بررسی چالش‌های برگردان نام سازمان‌ها نیز علت ب‌وجود آمدن چالش‌ها را بررسی کرده‌اند و به نتایجی مطابق با نتایج پژوهش حاضر دست یافته‌اند.

با توجه به هر یک از دلایل احتمالی پیش‌گفته و همچنین وضعیت کنونی برگردان نام‌های نویسندگان ایرانی در مهم‌ترین پایگاه استنادی جهان، و تنوع فراوان نام‌ها، جامعیت بازبایی نام‌ها می‌تواند تحت تأثیر قرار گیرد و تعداد زیادی از مدارک مرتبط هنگام جست‌وجوی کاربران به‌وسیله نام نویسندگان از دست رود. بنابراین، برای رفع این مشکل باید تمهیداتی برای کنترل یکدستی و بهنجارسازی نام‌ها در پایگاه در نظر گرفته شود تا در جست‌وجوی یک نام بتوان به حداکثر تولیدات علمی یک نویسنده دست پیدا کرد. بر این اساس، پیشنهاد می‌شود که دو سری راهکار در نظر گرفته شود: راهکارهای

موقت برای غلبه بر ناهنجاری‌های فعلی، که از جمله مهم‌ترین آن‌ها می‌توان به استفاده از عملگرها و نویسه‌های جایگزین بین صورت‌های نگارشی غالب، و ایجاد پیوندی بین آن‌ها اشاره کرد، و راهکار دائم که ایجاد ساندکس برای زبان فارسی و آواهای فارسی، مجهز کردن پایگاه به اصطلاحنامه و ارجاع دادن کاربر از صورت نگارشی پذیرفته‌نشده به صورت نگارشی پذیرفته‌شده و همچنین، تدوین جدول آوانگاری برای آواهای زبان فارسی بر اساس صورت‌های نگارشی غالب نویسندگان و توجه به اصل ساده‌نویسی در این الگوست. لازم است در پژوهش‌های آتی صورت‌های نگارشی نام چند نویسنده خاص را در چند پایگاه مختلف بررسی کرد و مشاهده کرد که آیا در همه پایگاه‌ها یک نام خاص به یک صورت نگارشی نمایه شده است یا نه. همچنین، با توجه به این که نتایج نشان داد که برای بهبود بازیابی، استانداردسازی ذخیره نام‌ها یک ضرورت است، پیشنهاد می‌شود که هنگام ذخیره نام‌ها در پایگاه، استاندارد تعیبه شود تا از چندگانگی نگارش نام‌ها جلوگیری شود.

فهرست منابع

- جهانشاهی، عبدالعلی. ۱۳۸۵. بررسی مشکلات دگرنویسی اسامی نویسندگان معاصر ایرانی. *اطلاعات‌شناسی* ۲ (۳): ۱۹۵-۲۱۴.
- جهانشیری، علی. ۱۳۹۸. نوشتن فارسی با الفبای لاتین. در ویلاگ تک نوشته‌ها: نوشتارهایی پیرامون زبان فارسی. <http://www.jahanshiri.ir/fa/fa/negaresh-be-latin>. (دسترسی در ۱۳۹۷/۶/۲۰).
- رائی ساریانقلی، محمدصابر. ۱۳۸۴. بررسی مشکلات جست‌وجو و بازیابی اطلاعات به زبان فارسی از اینترنت با مطالعه موردی بر روی کاربران مرکز اینترنت دانشگاه آزاد اسلامی واحد شبستر. پایان‌نامه کارشناسی ارشد، دانشگاه آزاد اسلامی، واحد تهران شمال.
- ریاحی‌نیا، نصرت، و معصومه نیک‌نیا. ۱۳۹۵. موانع به اشتراک‌گذاری مستندات اسامی اشخاص در فهرست مستند مجازی بین‌المللی (ویاف). *مطالعات ملی کتابداری و سازماندهی اطلاعات* ۲۷ (۳): ۵۷-۷۳.
- صادقی گورجی، شهربانو، علی‌اکبر پوراحمد، محسن حاجی‌زین‌العابدینی، و ثریا ضیایی. ۱۳۹۴. ارزیابی کارآمدی گوگل پژوهشگر در بازیابی اطلاعات نویسندگان دارای شکل‌های گوناگون نام: بررسی ضریب بازیافت و دقت. *پژوهشنامه کتابداری و اطلاع‌رسانی* ۵ (۱): ۲۰۵-۲۲۰.
- صدیق‌بهرادی، ماندانا. ۱۳۷۵. زبان فارسی در تبادل اطلاعات کتابشناختی. *نامه فرهنگستان* ۸: ۷۰-۸۲.
- _____. ۱۳۷۱. *ناهماهنگی ضبط نام‌های بیگانه در فارسی. فرهنگ* ۱۳: ۱۰۳-۱۱۸.
- عبدالهی نورعلی، محمدصادق. ۱۳۸۶. کندوکاو مسائل ریخت‌شناسی زبان فارسی در بازیابی اطلاعات از

- جست‌وجو گره‌های وب. پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، دانشگاه شیراز. شیراز. فلاحتی قدیمی فومنی، محمدرضا. ۱۳۹۲. *واژه‌نامه برگردان نام و نام خانوادگی نویسندگان خارجی (نوشته شده با حروف انگلیسی) به فارسی با استفاده از تحلیل رخدادهای محور*. شیراز: تخت جمشید.
- گل تاجی، مرضیه، و سعیده بذرگر. ۱۳۸۹. بررسی مشکلات ریخت‌شناسی زبان فارسی در سه پایگاه اطلاعاتی مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری، پژوهشگاه اطلاعات و مدارک علمی ایران و جهاد دانشگاهی. *فصلنامه کتابداری و اطلاع‌رسانی* ۱۳ (۲): ۱۹۱-۲۱۴.
- گل تاجی، مرضیه، و جواد عباس‌پور. ۱۳۹۶. بررسی عوامل واگرایی در نگارش وابستگی سازمان‌ها و مراکز پژوهشی و تأثیر آن بر جامعیت نتایج بازبایی شده در پایگاه تامسون رویترز. *فصلنامه کتابداری و اطلاع‌رسانی* ۲۰ (۱): ۸۸-۱۱۲.
- محقق‌زاده، محمدصادق و کاظم زارعیان. ۱۳۸۳. ارائه راه‌حل برای برخی مسائل اتوماسیون نگارش فارسی. *اطلاع‌رسانی* ۱۹ (۳-۴): ۱-۱۰.
- یارمحمدی، لطف‌الله. ۱۳۶۴. *درآمدی بر آواشناسی*. تهران: مرکز نشر دانشگاهی.

References

- Al-Onaizan, Y., K. & Knight. 2002. Machine transliteration of names in Arabic text. In *Proceedings of the ACL-02 workshop on Computational approaches to semitic languages*:1-13. Philadelphia, Pennsylvania, USA
- Deep, K., & V. Goyal. 2011. Development of a Punjabi to English transliteration system. *International Journal of Computer Science and Communication* 2 (2): 521-526.
- Falih, H. H. 2009. Some problems in the "translation" of English proper names into arabic. *Adab al-Basrah* 50: 42-54.
- Fumani, M. R. F. Q., M. Goltaji, & P. Parto. 2013. Inconsistent transliteration of Iranian university names: a hazard to Iran's ranking in ISI Web of Science. *Scientometrics* 95 (1): 371-384.
- Josan, G. S., & G. S. Lehal. 2010. A Punjabi to Hindi machine transliteration system. In *International Journal of Computational Linguistics & Chinese Language Processing* 15 (2): 77-102.
- Karimi, S., F. Scholer, & A. Turpin. 2007. Collapsed consonant and vowel models: New approaches for English-Persian transliteration and back-transliteration. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* 648-655 .. Prague, Czech Republic.
- Karoonboonyanan, T., V. Sornlertlamvanich, & S. Meknavin. 1997. A Thai Soundex system for spelling correction. In *Proceeding of the National Language Processing Pacific Rim Symposium* 633-636 .. Tokyo, Japan.
- Kawashima, H., & H. Tomizawa. 2015. Accuracy evaluation of Scopus Author ID based on the largest funding database in Japan. *Scientometrics* 103 (3): 1061-1071.
- Kay, B. N., & B. C. Rineer. 2012. Approaches to Arabic Name Transliteration and Matching in the DataFlux Quality Knowledge Base. In *The Fourth Workshop on Computational Approaches to Arabic Script-based Languages*. California, San diego.
- Kim, J., & J. Kim 2019 .. Effect of Forename String on Author Name Disambiguation. *Journal of the Association for Information Science and Technology* 1-17. <https://doi.org/10.1002/asi.24298>.
- Lerchenmueller, M. J., & O. Sorenson. 2016. Author disambiguation in PubMed: Evidence on the precision and recall of author-ity among NIH-funded scientists. *PLoS One* 11 (7): 1-13.

Yahia, M. E., R. E. A. Elhafez, T. B. Elsherif, & O. N. Osman. 2004. Intelligent soundex function for Arabic names. In *Proceedings. 2004 International Conference on Information and Communication Technologies: From Theory to Applications*, 407-408. Damascus, Syria.

Zahid, M. A., N. I. Rao, & A. M. Siddiqui. 2010. English to Urdu transliteration: An application of Soundex algorithm. In *2010 International Conference on Information and Emerging Technologies*. 1-5. Karachi, Pakistan.

مهسا کاوه

متولد سال ۱۳۷۱، دارای مدرک تحصیلی کارشناسی ارشد در رشته علم اطلاعات و دانش‌شناسی گرایش مدیریت اطلاعات از دانشگاه شیراز است.

بازیابی اطلاعات از جمله علایق پژوهشی وی است.



مهدیه میرزاییگی

متولد سال ۱۳۶۰، دارای مدرک تحصیلی دکتری در رشته علم اطلاعات و دانش‌شناسی از دانشگاه فردوسی مشهد است. ایشان هم‌اکنون دانشیار گروه علم اطلاعات و دانش‌شناسی دانشگاه شیراز است.

مباحث مرتبط با ربط در نظام‌های ذخیره و بازیابی اطلاعات، و عوامل شناختی تأثیرگذار بر فرایند بازیابی از جمله علایق پژوهشی وی است.



هاجر ستوده

دارای مدرک علم اطلاعات و دانش‌شناسی است. ایشان هم‌اکنون دانشیار دانشگاه شیراز است.

علم‌سنجی و دگرسنجی، سازماندهی و مدیریت دانش و بازیابی اطلاعات از جمله علایق پژوهشی وی است.



امیر سعید مولودی

متولد سال ۱۳۶۳، دارای مدرک تحصیلی دکتری در رشته زبان‌شناسی همگانی از دانشگاه تهران است. ایشان هم‌اکنون استادیار بخش زبان‌های خارجی و زبان‌شناسی دانشگاه شیراز است.

حوزه‌های زبان‌شناسی رایانه‌ای، زبان‌شناسی پیکره‌ای و زبان‌شناسی از جمله علایق پژوهشی وی است.

