

بررسی چالش‌ها و مسائل مرتبط به تایپ فارسی با استفاده از برنامه آوانگار

Google Transliteration

بهاره پهلوان‌زاده

استادیار پژوهشی طراحی و عملیات سیستم‌ها؛ مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری

فاطمه احمدی‌نسب

استادیار پژوهشی زبان‌شناسی رایانه‌ای؛ مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری

چکیده

در عصر حاضر با پیشرفت فناوری‌های رایانه‌ای بسیاری از امور توسط ماشین صورت می‌گیرد، که یکی از آنها تایپ یا حروف‌نگاری است که تا حد زیادی جای نوشتن با دست را گرفته است. خوشبختانه زبان فارسی یکی از زبان‌هایی است که دارای رسم‌الخط مرتبط با خود بوده و تا حدود زیادی با فضای رایانه و ماشین نیز مطابقت داده شده است. اما با وجود این برخی از فارسی‌زبانان داخل یا خارج از کشور به دلایل مختلفی مانند مهارت در تایپ لاتین، عدم توانایی در تایپ فارسی و یا عدم دسترسی به صفحه کلید فارسی ترجیح می‌دهند که از ابزارهای پینگلیش/ فینگلیش برای تایپ متون فارسی استفاده نمایند. در این زمینه تعداد بسیار محدودی ابزار تهیه شده که پرکاربردترین آنها ابزار آوانگار Google Transliteration است. نوشتار حاضر، در راستای این مهم به بررسی چالش‌ها و مسائل مرتبط به تایپ با این ابزار پرداخته است. این مسائل را می‌توان در ارتباط با رسم‌الخط فارسی و یا مسائل بین‌زبانی در ارتباط با تفاوت‌های دو خط لاتین و فارسی مطرح نمود که از این بین می‌توان به همزه، علائم زبرنجیری و کسره اضافه به عنوان برخی از چالش‌های مرتبط با رسم‌الخط فارسی و آوانگاری اسامی خارجی و تفاوت‌های نظام آوایی و نوشتاری دو زبان انگلیسی و فارسی به عنوان برخی از مسائل بین‌زبانی اشاره کرد. در این پژوهش پس از بررسی چالش‌های مذکور و موارد مرتبط دیگر به ارائه پیشنهادهایی برای بهبود این ابزار و افزایش دقت آن پرداخته می‌شود. علاوه بر این، پیشنهاداتی نیز در اختیار کاربران این نرم‌افزار قرار می‌گیرد تا با استفاده از آن بتوانند بر این مسائل فائق آیند.

کلیدواژه‌ها: ابزار پینگلیش/ فینگلیش، رسم‌الخط فارسی، ابزار آوانگار Google Transliteration، نظام نوشتاری فارسی، نظام نوشتاری لاتین.

مقدمه:

زبان فارسی جز زبان‌های هندواروپایی است که گستره وسیعی از گویشوران این زبان در کشورهای ایران، تاجیکستان و افغانستان متمرکز بوده و بعضاً در برخی از نواحی مختلف کشورهای آسیایی نیز به زبان فارسی تکلم می‌شود. علاوه بر این باید در نظر گرفت که تعداد زیادی از گویشوران فارسی‌زبان نیز در کشورهای اروپایی و آمریکایی پراکنده‌اند و همچنین تعداد زیادی از این افراد را دانشجویان فارسی‌زبان تشکیل می‌دهند. در این میان افراد فارسی‌زبانی وجود دارند که به دلایل مختلف مانند مهارت بیشتر در تایپ به خط لاتین، عدم دسترسی به صفحه کلید فارسی و یا حتی عدم آشنایی با رسم‌الخط فارسی (مانند مردم تاجیکستان، و نسل دوم مهاجران) به ناچار ترجیح می‌دهند برای تولید ماشینی متون فارسی از ابزار پینگلیش/ فینگلیش استفاده نمایند. این در حالی است که کاربر در استفاده از این ابزار با مسائل و چالش‌های نوشتاری روبرو خواهد شد که از ویژگی‌های منحصر به فرد رسم‌الخط فارسی و یا از مسائل بین‌زبانی دو زبان انگلیسی و فارسی نشئت می‌گیرد. قبل از پرداختن به مبحث پژوهش لازم است که به اختصار به معرفی ابزار Google Transliteration و چرایی انتخاب آن پرداخته شود. بررسی‌های مختلف نگارندگان این نوشتار نشان داد که تعداد اندکی ابزار در این حوزه وجود دارد که بیشتر آنها بسیار جدید بوده و علاوه بر این از ابزار فینگلیش گوگل الگوبرداری کرده‌اند. با توجه به محبوبیت جهانی، گستره کاربردها، قابلیت‌های ابزارها و چندزبانگی گوگل، نیازسنجی و به‌روزرسانی مداوم امکانات آن و بالاخره مراجعه بیشتر پژوهشگران حوزه‌های مختلف دانش برای استفاده از خدمات متعدد این سرویس دهنده جهانی در این پژوهش به مطالعه این ابزار پرداخته می‌شود.

معرفی برنامه آوانگار Google Transliteration :

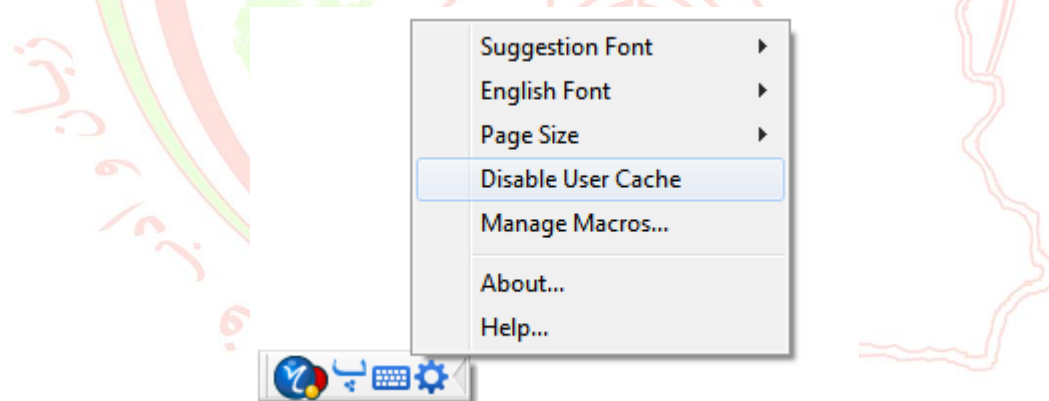
گوگل برای ویندوز اخیراً ابزار جدیدی تحت عنوان آوانگاری خودکار (Google Transliteration) ارائه نموده است. این ابزار آوانگار از صفحه Google Input Tools on Windows (لینک <https://www.google.com/inputtools/windows/index.html>) قابل دسترسی می‌باشد. این ابزار در حال حاضر برای ۲۲ زبان مانند عربی، فارسی، یونانی، عبری قابل استفاده است.

کاربر به وسیله این ابزار قادر به تبدیل نویسه‌های لاتین به زبان مدنظر می‌باشد و بدین‌وسیله با وارد کردن آواهای یک زبان به حالت رسم‌الخط لاتین (مانند فینگلیش) قادر به دریافت خروجی این ابزار به رسم‌الخط مد نظر (مانند فارسی) خواهد بود. باید توجه نمود که رفتار این ابزار آوانگار (Transliteration) کاملاً متفاوت از سایر ابزارها یا نرم افزارهای مترجم (Translator) است؛ و در واقع این ابزار آوانگار همان‌طور که از نامش گویاست مبدل آوای هر واژه از الفبای یک زبان به الفبای زبان دیگر است نه ترجمه معنایی آن.

ویژگی‌های برنامه آوانگار Google Transliteration:

فعال بودن این ابزار بر خروجی صفحه کلید و هر نرم افزار ویرایشگر متن اثرگذار خواهد بود، به این نحو که برای مثال نویسه‌های یک واژه فینگلیش تا پایان یافتن واژه یکی پس از دیگری به فارسی تبدیل می‌شود و به محض وارد کردن فاصله یا علائم نقطه-گذاری خروجی به شکل یک واژه در رسم‌الخط فارسی به نمایش در می‌آید. به برخی از مزایای این ابزار به صورت فهرست‌وار در ذیل اشاره می‌شود:

- پشتیبانی آفلاین: بدون دسترسی به اینترنت این ابزار همچنان قادر به کار می‌باشد.
- تکمیل واژه: این ابزار بر مبنای حروف آغازین واژه با جستجو در فرهنگ لغت قادر به تکمیل واژه می‌باشد.
- شخصی‌سازی انتخاب‌ها: این ابزار بر مبنای تصحیحات قبلی متناسب با قوانین دستوری کاربر، قادر به پیشنهاد واژه‌های صحیح می‌باشد. به بیان دیگر این ابزار امکان قرار گرفتن تحت آموزش توسط کاربر را دارد.
- آخرین انتخاب: این ابزار قادر به شخصی‌سازی و حفظ آخرین انتخاب واژه پیشنهادی کاربر - متناسب با اکانت کاربری بر کلیه برنامه‌های کاربردی - است. باید توجه داشت که این شخصی‌سازی فقط جهت تکمیل واژه‌های مشخص شده در سه انتخاب اول می‌باشد که در صورت صلاحدید کاربر به روش زیر می‌تواند غیر فعال گردد.



- صفحه کلید آسان: این ابزار بر مبنای فرهنگ لغت قادر به پیشنهاد واژه‌های پیچیده و کم‌بسامد می‌باشد.
- سفارشی‌سازی: اندازه پنجره و فونتها با توجه به نظر کاربر قابلیت تغییر و سفارشی‌سازی دارد.
- جستجوی سریع: با یک تک کلیک بر روی واژه‌های لایت شده قادر به جستجو تحت وب می‌باشد.
- تغییر صفحه کلید چندزبانه: این ابزار به راحتی قادر به خاموش - روشن شدن جهت سوئیچ کردن به خط لاتین در یک متن ترکیبی می‌باشد.

- راهنما: این ابزار دارای راهنمای آسان و کارآمد به همراه کلیه میان‌برهای کاربردی است.

پیشینه مطالعات:

با جستجویی که توسط نگارندگان نوشتار حاضر صورت گرفت، به نظر می‌رسد که در مورد تایپ فارسی با استفاده از ابزار فینگلیش گوگل به طور اخص مطالعه و پژوهش خاصی صورت نگرفته است، اما پژوهشگران حوزه‌های مختلف به فراخور دغدغه‌ها و زمینه‌های تخصصی خود نسبت به خط فارسی، پژوهش‌هایی را در این مورد انجام داده‌اند که در ذیل به برخی از آنها اشاره می‌شود. آشوری (۱۳۷۵) با اشاره به مشکلات شناسه‌ها در زبان فارسی، برای اصلاح خط فارسی پیشنهاد می‌دهد که صورت‌های صرفی زمان حال فعل «بودن» و ضمائر متصل، به صورت جدا و منفصل از کلمه‌های قبل از خود نوشته شوند. معصومی همدانی (۱۳۸۱) در مقاله‌ای تحت عنوان «خط فارسی و رایانه»، با اشاره به ضرورت یکسان‌سازی رسم‌الخط فارسی به منظور استفاده از خطایاب، امکان جستجوی واژه در متن و تهیه نمایه، راه‌حل را در تهیه یک دستورالعمل واحد برای خط فارسی می‌داند. اسلامی (۱۳۸۱) وضع یک نشانه اصلی برای نمایش کسره اضافه، استفاده از نشانه‌های متفاوت برای نمایش «ی» نکره، «ی» اسم‌ساز و «ی» صفت‌ساز، سرهم-نویسی کلمات غیربسیط، قراردادن علامت‌های جداگانه برای واژه‌بست‌های ربطی فعل «بودن» و ضمائر ملکی، و بالاخره نشان دادن اسامی خاص را در جهت اصلاح خط فارسی پیشنهاد می‌کند. محقق‌زاده و زارعیان (۱۳۸۳) چپ‌نویسی، کاهش حروف از طریق جدانویسی، قائل شدن به دو شکل کوچک و بزرگ حروف را در مورد زبان فارسی پیشنهاد می‌کنند. علاوه بر این، نگارش حروفی که خواننده شده ولی نوشته نمی‌شوند، ایجاد علامتی برای کسره اضافه و عدم تمایز بین «آ» و «ا» را نیز در جهت سازگار ساختن زبان فارسی با محیط‌های رایانه‌ای ضروری می‌دانند. واضح است که این دو پژوهشگر غالب ویژگی‌های خط لاتین را الگوی پیشنهادات خود قرار داده‌اند. اسلامی (۱۳۸۶) نقطه‌دار بودن برخی از نویسه‌های فارسی را یکی از ایرادات خط فارسی می‌داند، چرا که علاوه بر دشوار نمودن املای کلمات، این ایراد اساسی را دارد که استفاده از برنامه نشانه‌خوان نوری را برای متون فارسی دشوار می‌کند. گل-تاجی و بذرگر (۱۳۸۹) مشکلات ریخت‌شناسی خط فارسی را در سه پایگاه اطلاعاتی مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری، پژوهشگاه اطلاعات و مدارک علمی ایران، و جهاد دانشگاهی با یکدیگر مقایسه کرده و نتیجه می‌گیرند که هیچ‌کدام از پایگاه‌های مذکور به مسائل خط فارسی توجه نداشته‌اند. صفارمقدم (۱۳۸۶) فاصله‌گذاری را در ترکیبات فارسی ضروری دانسته و مفضلاً به مبحث فاصله‌گذاری درون کلمه‌ای و برون کلمه‌ای در انواع کلمات فارسی پرداخته است.

همانطور که از مرور پیشینه واضح است، آن است که در مورد خط فارسی در بافت‌های مختلف پژوهش‌های متعددی صورت گرفته است. اما به نظر می‌رسد که علی‌رغم وجود گوگل به عنوان یک سرویس دهنده جهانی موفق و فراگیر که دارای امکانات و ابزارهای مختلفی است و زبان فارسی را نیز پوشش می‌دهد، به مطالعه مسائل زبان فارسی در این محیط توجه چندانی نشده است. لذا در راستای پر کردن این خلأ مطالعاتی - هر چند در مقیاسی کوچک - نوشتار حاضر با توجه به اهمیت و ضرورت ارتقای خط فارسی در حوزه فناوری‌های نوین به صورت اعم و ابزار فینگلیش به طور اخص، به بررسی مسائل خط و زبان فارسی در این ابزار پرداخته است.

روش پژوهش:

به منظور انجام پژوهش از روش درون‌دهی نمونه متون استفاده شد. در این روش، متون مختلف فارسی مانند متون علمی، خبری، محاوره‌ای و داستانی بصورت فینگلیش وارد ابزار شده و سپس متن فارسی حاصل از جهت مطابقت با رسم‌الخط زبان فارسی بررسی گردید. لازم به ذکر است که تایپ در این ابزار به بافت و متن حساس نبوده و واژه-بنیاد است. بنابراین چالش‌های مورد نظر تایپ فینگلیش در سطح واژه مورد آزمایش و بررسی قرار گرفت.

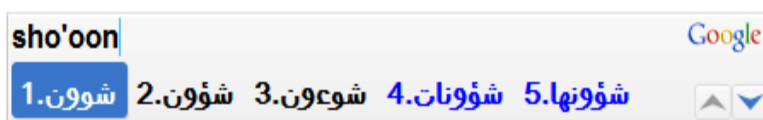
یافته‌های پژوهش:

با بررسی متون فارسی حاصل از تایپ فینگلیش، مسائل و چالش‌های استفاده از ابزار به صورت زیر و تحت عناوین مختلفی گردآوری شد که ذیل هر کدام با ذکر مثال به بحث پیرامون آن پرداخته می‌شود:

۱. همزه: در زبان فارسی برخی از واژه‌های قرصی که از زبان عربی وارد فارسی شده‌اند دارای نویسه همزه هستند. همزه یکی از نویسه‌هایی است که با توجه به واژه به شکلهای متفاوتی نوشته می‌شود برای مثال مؤید، مسئله، تألیف، شیء و... یکی از مشکلات نگارشی ابزار فینگلیش گوگل عدم استفاده از علامت صحیح همزه در واژه است. برای مثال با تایپ moayyed در ابزار فینگلیش، ابزار گزینه‌های زیر را پیشنهاد می‌دهد:



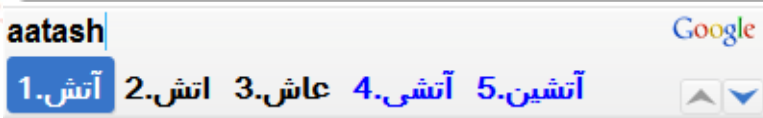
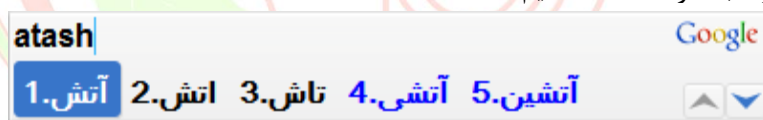
تجربه استفاده از ابزار نشان می‌دهد که استفاده از کلید «'» برای آوانگاری همزه تا حدود زیادی مشکل‌گشا است. اما با این وجود در مورد بسیاری از واژه‌های دارای علامت همزه، املاي نادرست را پیشنهاد می‌دهد. برای مثال با تایپ sho'oon نتایج زیر به دست آمد:



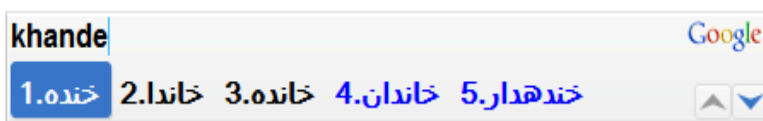
بنابراین یکی از مواردی که در اصلاح ابزار باید مورد توجه قرار گیرد به واژگان دارای علامت همزه مرتبط می‌شود.
۲. علامت ع: از دشواری‌های دیگر استفاده از ابزار فینگلیش می‌توان به تایپ کلمات دارای نویسه «ع» اشاره کرد. از آنجا که این نویسه معادلی در زبان انگلیسی ندارد به اشکال مختلفی آوانگاری می‌شود که خود این امر به تفاوت و تشتت در نگارش و در نتیجه عدم دقت در واژه فارسی منجر می‌شود. برای مثال برای تایپ فینگلیش این نویسه، از شیوه‌های مختلفی مانند استفاده از aa, a', ea و حتی عدم نگارش آن مخصوصاً در جایگاه آغازین واژه اشاره کرد. در زیر مثالی از این نویسه آورده شده است:



مشکلاتی از این دست سبب شده است که با تایپ atash یا aatash نتوانیم به املاي واژه «عطش» دست یابیم مگر اینکه به جای تایپ واژه کامل از تایپ حرف به حرف استفاده کنیم.

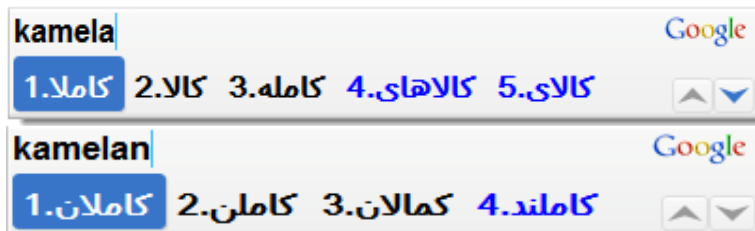


پیشنهاد می‌شود که ابزار فینگلیش این نوع کلمات فارسی را که دارای صورت فینگلیش یکسان هستند تحت عنوان واژگان دارای ابهام نویسه‌ای فینگلیش دسته‌بندی کند و در گزینه‌های پیشنهادی هر دو واژه را پیشنهاد دهد. برای مثال برای افزایش کارایی ابزار فینگلیش، در مورد واژه khande باید دو گزینه «خنده» و «خوانده» پیشنهاد شود تا کاربر با سهولت بیشتری ابزار را مورد استفاده قرار دهد.



۳. علامت زبرزنجیری، تشدید، تنوین: در ابزار فینگلیش امکان نمایش علائم زبرزنجیری مانند تنوین، تشدید و «ی» کوتاه وجود ندارد. این در حالی است که در مواردی به نمایش این علائم نیاز است. در مورد تایپ کلمات تنوین‌دار مانند کاملاً، غالباً

و دهها کلمه تنوین دار دیگر، تایپ فینگلیش به واژه‌هایی مانند کاملن، غالبن و... منجر می‌شود که املائی درست واژه نیست. در مورد اخیر یا باید فینگلیش واژه بدون تایپ نویسه n انجام گیرد که مغایر شمّ زبانی فارسی‌زبانان است و یا آنکه بعد از تایپ واژه کامل «ن» واژه حذف شود.

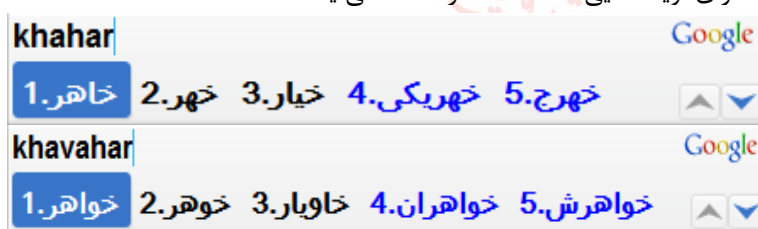


۴. **واکه‌ها:** همانطور که می‌دانیم در زبان فارسی واکه‌ها به دو نوع کوتاه و بلند تقسیم‌بندی می‌شوند و در مجموع سه واکه کوتاه /a/, /e/, /o/ و سه واکه بلند /a:/, /e:/, /u:/ وجود دارد که واکه‌های بلند به صورت نویسه در املائی واژه نشان داده می‌شوند اما واکه‌های کوتاه بجز در جایگاه آغازین واژه غالباً در املائی کلمات از نویسه خاصی برخوردار نیستند. با توجه به این نکته که غالباً نمود نوشتاری واکه‌های فارسی منظم نبوده و از پیچیدگی بالایی برخوردار است، طبیعی است که با تایپ فینگلیش واژگانی نیز پیشنهاد خواهد شد که از واژه مورد نظر کاربر دور خواهد بود. این در حالی است که سلیقه‌های متنوع کاربران ابزار فینگلیش این پیچیدگی را افزایش می‌دهد. برای مثال برخی از افراد برای نمایش واکه /u:/ از /ou, oo, u/ و حتی O استفاده می‌کنند که این امر خود بر پیچیدگی نمایش واکه‌ها می‌افزاید. به نظر می‌رسد که در این خصوص با مطالعه گسترده تمایلات نگارشی کاربران در نحوه استفاده از حروف انگلیسی برای نمایش واکه‌های فارسی و یا قاعده‌مند کردن نمایش این واکه‌ها توسط یک دستورالعمل، می‌توان دقت ابزار مورد بحث را افزایش داد.

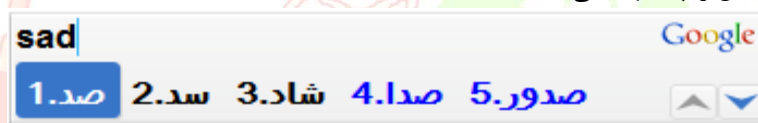
۵. **ی میانجی:** در زبان فارسی هم‌نشینی واکه‌ها امکان‌پذیر نیست و برای جلوگیری از این اتفاق از همخوان میانجی مانند «ی» استفاده می‌شود که در تلفظ و در برخی از موارد در خط نیز نمود پیدا می‌کند. با تایپ فینگلیش واژه‌های دارای واج میانجی، می‌توان دید که یکی از دشواری‌های تایپ فینگلیش به واج میانجی مربوط می‌شود، چرا که به نادرست شدن املائی واژه منجر می‌شود. برای مثال با تایپ واژه didgaha برنامه فینگلیش به درستی واژه «دیدگاهها» را در اختیار کاربر قرار می‌دهد. اما با تایپ didgahaye واژه‌های نادرستی پیشنهاد می‌شود. مثال دیگر به واژه darbare مربوط می‌شود که غالباً به صورت «درباره» تلفظ می‌شود که با تایپ darbareye واژه «درباری» و چند واژه نامرتبط دیگر پیشنهاد می‌شود. با توجه به این موضوع، پیشنهاد می‌شود که وضعیت نگارش «ی» میانجی در دستور کار برنامه‌نویسان برنامه فینگلیش گوگل قرار گیرد. هرچند که کاربر با تجربه کار با این برنامه به این نتیجه خواهد رسید که ابتدا واژه را تایپ کند و سپس در ویرایش خود شخصاً «ی» میانجی را تایپ نماید. البته پرواضح است که بهینه‌سازی برنامه فینگلیش بر واگذاری تصحیح به کاربر ارجح است.



۶. نویسه‌هایی که نوشته می‌شوند اما خوانده نمی‌شوند: در زبان فارسی واژه‌هایی وجود دارند که برخی از نویسه‌های آنها خوانده نمی‌شود. این امر به علت تغییر تاریخی زبان و محافظه‌کار بودن خط است. برای مثال «و» در واژه «خواهر» در گذشته‌های دور زبان فارسی تلفظ می‌شده است و هنوز هم در برخی از گویش‌ها تلفظ می‌شود. اما در فارسی معیار کاملاً ناپدید شده است و تنها در خط دیده می‌شود. از واژه‌های «خواب»، «خوار»، «خویش» می‌توان به عنوان مثالهای دیگر یاد نمود. با تایپ فینگلیش این واژه‌ها می‌توان دید که برنامه واژه‌های نادرستی را در اختیار کاربر قرار می‌دهد مگر آنکه کاربر نویسه بی‌آوا را در تایپ فینگلیش خود لحاظ کند که این امر با شم زبانی در تضاد هست و یا آنکه ممکن است کاربر فارسی‌زبان ناآشنا به خط از وجود چنین نویسه‌ای در واژه مورد نظر آگاه نباشد. بنابراین یکی از مسائل دیگر که باید مورد توجه برنامه‌نویسان این برنامه قرار گیرد، به واژه‌هایی مربوط می‌شود که دارای نویسه‌هایی هستند که در تلفظ نمی‌آیند.



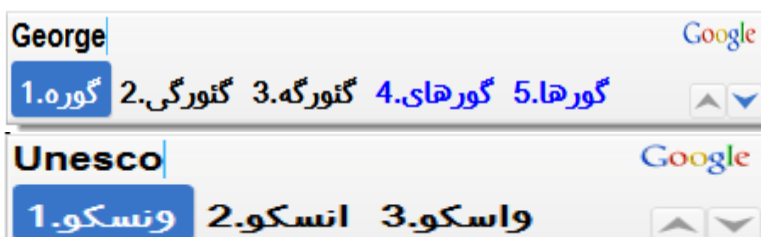
۷. هم‌آوایی: برخی از واژه‌های زبان فارسی هم‌آوا هستند. این بدان معناست که علی‌رغم تفاوت در صورتهای نگارشی، یکسان تلفظ می‌شوند که این امر باعث می‌شود هنگام تایپ فینگلیش، ابهام نویسه‌ای فینگلیش روی دهد مانند واژه‌های خا/خوار، خویش/خویش، صد/صد، ارز/عرض، خورد/خرد و مثالهای دیگر. در مورد برخی از این مثالها مانند صد/صد و خورد/خرد، ابزار فینگلیش هر دو صورت محتمل را پیشنهاد می‌دهد.



لازم است که این امکان در مورد مثالهای دیگر ذکر شده در این بند و البته واژه‌های دیگر هم‌آوای فارسی نیز فراهم شود.

۸. فاصله‌گذاری - پیوسته‌نویسی و جدانویسی کلمات مرکب و مشتق: در زبان فارسی برخی از حروف با توجه به جایگاهشان در واژه، به صورت جدا و یا پیوسته با حروف دیگر نوشته می‌شوند که در تایپ فارسی فاصله‌گذاری بر حسب مورد با یک فاصله و یا نیم‌فاصله نمایش داده می‌شود. در ابزار فینگلیش امکان نمایش فاصله وجود ندارد. برای مثال با تایپ *dirinshenasi* واژه «دیرینشناسی» به دست می‌آید که باید به صورت «دیرین‌شناسی» نمایش داده شود. مثال دیگر به تایپ فینگلیش *neshaneshenakhti/neshanehshenakhti* مرتبط می‌شود. با تایپ این دو صورت، واژه «نشان‌شناختی» به دست داده می‌شود که املا نادرستی از واژه است و کاربر برای رسیدن به املا درست باید *shenakhti* را به صورت واژه‌ای جدا تایپ کند که این امر با شم زبانی فارسی‌زبان در تضاد است. متأسفانه امکان فاصله‌گذاری در ابزار فینگلیش وجود ندارد و با استفاده از کلید *space* واژه در محیط ابزار تایپ می‌شود. به نظر می‌رسد که به دلیل نیاز به فاصله‌گذاری در درون اجزاء جمله لازم است که برای این مورد چاره‌اندیشی شود تا امکان استفاده از نیم‌فاصله در واژگان مرکب و یا مشتق فراهم شود.

۹. آوانگاری اسامی خارجی شامل اسامی خاص افراد، سازمان‌ها و کشورها: با تایپ اسامی خارجی شامل اسامی خاص، شامل اسامی افراد، مکانها و سازمانها در برنامه فینگلیش، با آنها مانند واژگان عادی برخورد می‌شود به طوری که برای اسامی و نامهای مشهوری حتی مانند جرج، یونسکو و... گزینه‌هایی غیرمحتمل پیشنهاد می‌شود. یکی از راهکارها برای بهبود برنامه فینگلیش تجهیز آن به فرهنگ اسامی و نامهای خاص است تا به وسیله آن امکان آوانگاری صحیح اسامی خاص و حتی تبدیل آن به معادل رایج فارسی آن فراهم شود. برای مثال لازم است که با ارتقای برنامه فینگلیش امکانی فراهم شود که مثلاً با تایپ *Netherland* واژه «هلند» پیشنهاد شود.



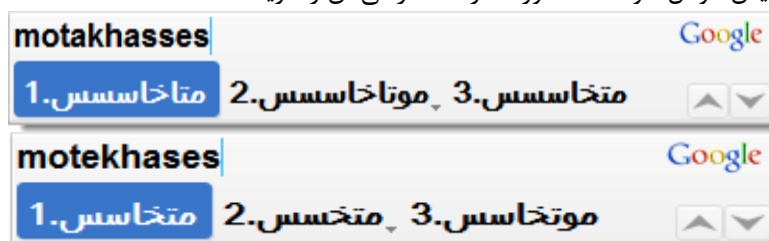
۱۰. **نویسه‌های ترکیبی:** در فارسی تعدادی واج وجود دارد که در انگلیسی معادل نداشته و غالباً به صورت ترکیبی از دو نویسه به نمایش گذارده می‌شوند مانند gh به جای «ق»، zh به جای «ژ» و kh به جای «خ». اما با این حال برخی از افراد همین سه واج را به ترتیب با سه نویسه q, j و x نشان می‌دهند. در ضمن «غ» نیز غالباً با gh و q نشان داده می‌شود. این در حالی است که در برخی از موارد برنامه‌های فینگلیش این دو گزینه محتمل را پیش‌بینی نکرده و متناسب با آن پیشنهاداتی را ارائه نکرده است. برای مثال با تایپ ghorooob در برنامه فینگلیش به درستی «غروب» پیشنهاد می‌شود. اما با تایپ qoroob واژه مرتبطی پیشنهاد نمی‌شود. با بررسی تمایل غالب فارسی‌زبانان در استفاده از فینگلیش برای به نمایش گذاردن نویسه‌های ترکیبی، می‌توان فهرستی از نگارش‌های مختلف و محتمل این واج‌ها را گردآوری کرد تا برنامه فینگلیش بتواند بر اساس آن واژه‌های ممکن را پیشنهاد دهد. البته این امکان در مورد نگارش متعدد و مختلف واژه‌ها در محیط برنامه وجود دارد که یکی از نقاط قوت برنامه فینگلیش به شمار می‌رود. برای مثال با تایپ هر پنج صورت فینگلیش rowshan, roshan, rooshan, roushan, rushan اولین گزینه پیشنهادی همان واژه «روشن» است. اما ذکر این نکته ضروری است که برنامه فینگلیش از این مزیت نیز برخوردار است که تا حدود زیادی هم‌نشینی‌های اتفاقی را از صداهای ترکیبی تشخیص داده و واژه درست را پیشنهاد می‌دهد برای مثال با تایپ ashab و eshal به درستی واژه‌های «اصحاب» و «اسهال» در اختیار کاربر قرار می‌گیرد.



۱۱. **واژه‌های دارای غلط‌های مصطلح:** در زبان فارسی تعدادی واژه وجود دارد که به دلایل مختلف از جمله وام‌گیری از زبان عربی، داشتن معنای متشابه با واژه‌های دیگر و یا دلایل تاریخی مختلف دارای نگارش‌های نادرست رایج نیز هستند مانند واژه‌های «سیاست‌گذاری» و «نعناع» که املاهای دیگر آنها یعنی «سیاست‌گزاری» و «نعنا» نیز به کار می‌رود. در این خصوص پیشنهاد می‌شود که در برنامه فینگلیش امکانی برای نمایش املاهای درست واژه از املاهای نادرست فراهم شود تا کاربر به استفاده از املا درست ترغیب شود. در ضمن پیشنهاد می‌گردد که واژه دارای املا درست پیشنهاد اول و غلط‌های مصطلح جز پیشنهادهای بعدی فینگلیش باشد که اکنون در برخی از موارد این ترتیب رعایت نمی‌شود برای مثال برنامه فینگلیش ابتدا واژه «نعنا» و سپس «نعناع» را پیشنهاد می‌دهد.



۱۲. **وام‌واژه‌های عربی دارای بیشترین حروف خاص عربی:** به نظر می‌رسد که دشوارترین واژه‌ها برای تایپ فینگیلیش واژه‌های فرضی از زبان عربی هستند که دارای چند حرف خاص عربی در یک واژه هستند مانند واژه‌های «متخصص» و «اصطلاح». این در حالی است که این واژه‌ها در زبان فارسی کاملاً رایج هستند. نویسندگان نوشتار حاضر با ورود صورتهای مختلف فینگیلیش این دو واژه در برنامه فینگیلیش گوگل نتوانستند صورت درست فارسی آن را دریافت کنند.



۱۳. **وام‌واژه‌های دارای «ال»:** برخی از واژگان فارسی وام‌واژه‌هایی هستند که از زبان عربی وارد فارسی شده‌اند و دارای «ال» هستند که یک عنصر معرفه‌ساز است. تلفظ این واژه‌ها غالباً با املای واژه متفاوت است؛ برای مثال «ال» واژه دارای «ال» در حالت مضاف‌الیه غالباً تلفظ نمی‌شود. این امر باعث می‌شود که در تایپ فینگیلیش این واژه‌ها، املاهایی نادرست نیز برای واژه فینگیلیش پیشنهاد شود برای مثال در مورد واژه‌های «بالعکس»، «بالاخره» و «بالذات» املاهای «بلعکس»، «بلاخره» و «بذات» «بذات»، «بلاخره» و «بلعکس» پیشنهاد می‌شود و املاهای صحیح یا اصلاً پیشنهاد نمی‌گردند و یا در صورت پیشنهاد جزء اولویت‌های بعدی هستند. توجه به این واژه‌ها نیز باید در دستور کار برنامه‌نویسان این برنامه قرار بگیرد.



۱۴. **تأثیر تلفظ‌های مختلف واژه بر ترتیب واژه‌های پیشنهادی:** برخی از واژه‌های فارسی به دلایل مختلف از جمله دلایل کاربردشناختی، دارای تلفظ‌های مختلفی هستند مانند واژه‌های محبت/مُحبت، عدالت/عِدالت، عطر/عِطر، عِدوت/عِدوت، مَداد/مِداد و... که قاعدتاً تایپ فینگیلیش این واژه‌ها با هر کدام از تلفظ‌های مورد نظر بر واژه‌های پیشنهادی برنامه تأثیرگذار خواهد بود. اما با تایپ فینگیلیش این تلفظ‌های مختلف یا واژه مورد نظر اصلاً پیشنهاد نمی‌شود و یا در صورت پیشنهاد اولویت آنها تغییر می‌کند. برای مثال با تایپ atr واژه «عطر» و با تایپ etr واژه «اتر» و سپس «عطر» پیشنهاد می‌شود. مثال دیگر به mohebbat مربوط می‌شود که به پیشنهاد واژه «محبت» و mohabbat به پیشنهاد واژه «موهبت» به عنوان اولویت اول منجر می‌شود.





۱۵. ه غیرملفوظ پایان واژه: در زبان فارسی برخی از واژه‌ها به ه ختم می‌شود که تلفظ نمی‌شود مانند «خانه»، «پله»، «زمانه»، «عمده» و... که این واقعیت باعث شده است که کاربران فینگلیش در آوانگاری این واژه‌ها در تایپ «ه» غیرملفوظ مختار باشند. برای مثال «خانه» را می‌توان به دو صورت khaneh و khane آوانگاری نمود. بررسی برنامه فینگلیش نشان می‌دهد که ظاهراً این قابلیت در برنامه لحاظ شده است. علاوه بر این، در زبان فارسی از کسره اضافه که غالباً در خط فارسی نشان داده نمی‌شود برای بیان رابطه اضافی (مضاف- مضاف‌الیه) استفاده می‌شود. این امر باعث می‌شود که بسیاری از واژه‌های فارسی مختوم به «ه» غیرملفوظ با برخی از واژه‌ها در حالت اضافی هم‌آوا شوند مانند واژه‌های رابط/ رابطه، بهار/ بهاره، مهر/ مهره، بال/ باله، مال/ ماله و منظر/ منظره. یکی از راه‌حل‌های جلوگیری از این نوع ابهام، عدم استفاده کاربران از کسره اضافه در تایپ فینگلیش است. اگر چه با تایپ omde در برنامه فینگلیش با پیشنهاد واژه «عمده» به عنوان اولویت اول و با تایپ omdeh با پیشنهاد واژه «اومده» به عنوان اولویت اول روبرو می‌شویم، حال آنکه با تایپ هر دو صورت فینگلیش انتظار داریم که واژه «عمده» را دریافت کنیم. توجه به این موضوع به عنوان یکی از چالش‌های زبان فارسی در برنامه فینگلیش باید مورد توجه قرار گیرد.



۱۶. واحدهای بزرگتر از واژه: در زبان فارسی مانند اغلب زبانها امکان تجزیه جمله به واحدهای نحوی وجود دارد که این واحدها شامل گروه اسمی، گروه فعلی، گروه حرف اضافه‌ای و... هستند. شمّ زبانی اقتضا می‌کند که برنامه فینگلیش گوگل در حال حاضر و با توجه به میزان فناوری به کار رفته در آن، واحد واژه و یا نهایتاً گروه دستوری مانند گروه اسمی را پیشنهاد دهد. در بسیاری از موارد، پیشنهاد‌های برنامه فینگلیش گوگل حاکی از آن است که این برنامه پا را از واژه فراتر گذارده است. اگر چه این پیشرفت قابل تحسین است اما واحدهای زبانی پیشنهاد شده غالباً به گروه اسمی مضاف و مضاف‌الیه محدود می‌شود که مضاف‌الیه را ضمائر ملکی و مفعولی و یا صورتهای متصل فعلی فعل «بودن» تشکیل می‌دهد. برای مثال با تایپ ezdevaj در برنامه فینگلیش پیشنهادات زیر در اختیار کاربر قرار می‌گیرد.



در برنامه فینگلیش امکان استفاده از فاصله وجود ندارد و به محض استفاده از این کلید، واژه فارسی مورد نظر تایپ شده و برنامه منتظر تایپ واژه‌های دیگر از طرف کاربر باقی می‌ماند. به نظر می‌رسد علت محدود شدن برنامه فینگلیش به این نوع واژه‌ها دیده شدن اسم+ ضمیر متصل ملکی به عنوان یک واژه باشد. با ارتقاء برنامه فینگلیش و ایجاد امکان استفاده از فاصله به عنوان یک کاراکتر درون گروه اسمی بتوان امکان پیشنهاد باهم‌آیی‌ها را فراهم نمود. باهم‌آیی‌ها را می‌توان بر اساس ترکیبات واژگانی پربسامد به دست آورد.

بحث و نتیجه‌گیری:

با توجه به اهمیت و ضرورت ارتقای خط فارسی در محیط‌های رایانه‌ای، در این نوشتار به معرفی یکی از ابزارهای جدید پیشنهادی سرویس دهنده جهانی گوگل با عنوان برنامه آوانگار Google Transliteration و همچنین بررسی چالشها و مسائل مرتبط به تایپ با این ابزار پرداخته شد.

چالشها و مسائل ابزار فینگلیش در قالب شانزده مورد به تفصیل در متن نوشتار مورد بحث و بررسی قرار گرفت. در نهایت به اختصار می‌توان اینگونه بیان کرد که در مورد واژه‌های قرضی با نویسه‌های خاص (همزه، علامت زبر زنجیری، تشدید، تنوین، «ی» کوتاه و یا واکه‌های کوتاه /a/, /e/, /o/ و یا سایر نویسه‌هایی که معادل آنها در زبان انگلیسی نیست یا کم‌بسامد می‌باشد (برای مثال نویسه «ع» یا «gh») و بنابراین منجر به آوانگاری‌های مختلفی می‌شود؛ می‌توان با اعمال دستورات عملیاتی یکپارچه برای کاربران یا گردآوری فهرستی از نگارش‌های مختلف و محتمل واژه‌ها در قالب یک پیکره یا فرهنگ لغات برای این ابزار، به امکان پیشنهاد واژه‌های مرتبط‌تر توسط ابزار مذکور کمک شایانی کرد و بدینوسیله دقت آن را افزایش داد. در ضمن برای بهبود برنامه فینگلیش تجهیز کردن آن به فرهنگ اسامی و نامهای خاص افراد، سازمان‌ها و کشورها پیشنهاد می‌گردد تا به وسیله آن امکان آوانگاری صحیح اسامی خاص و حتی تبدیل آن به برابر نهاده رایج فارسی فراهم شود. علاوه بر این، در مورد واژه‌های هم‌آوا نیز بهتر است کلیه موارد گردآوری شود و ابزار فینگلیش هر دو صورت محتمل را پیشنهاد دهد. از طرفی، در مورد وام‌واژه‌هایی که کاربر ناگزیر به تایپ حرف به حرف آن به جای تایپ واژه کامل است، برای افزایش کارآیی پیشنهاد می‌شود که ابزار فینگلیش گوگل این نوع کلمات فارسی را تحت عنوان واژگان دارای ابهام نویسه‌ای فینگلیش دسته‌بندی و در گزینه‌های پیشنهادی هر دو واژه را پیشنهاد دهد تا کاربر با سهولت بیشتری ابزار را مورد استفاده قرار دهد.

همچنین به دلیل پیچیدگی‌های خاص زبان فارسی، یکی از عمده نقاط چالش برانگیز مقوله فاصله‌گذاری یعنی پیوسته‌نویسی و جدانویسی کلمات مرکب و مشتق است که همانطور که در بند ۸ مطرح شد در ابزار فینگلیش نیز به خوبی تعبیه نگشته است که این مهم نیاز به بازبینی توسط پژوهشگران زبان‌شناس و علوم رایانه دارد و همچنین پیشنهاد می‌گردد جهت بهینه‌سازی این ابزار وضعیت نگارش «ی» میانجی در دستور کار برنامه‌نویسان برنامه فینگلیش گوگل قرار گیرد. مسلماً با توجه به پیشنهادات فوق، می‌توان با ارائه و اجرای راه‌حل‌های زبان‌شناختی و رایانه‌ای وضعیت ابزار مورد بحث را ارتقاء داد تا از این راه به گسترش و بهبود وضعیت زبان فارسی در محیط‌های رایانه‌ای کمک نمود.

فهرست منابع:

- آشوری، د. (۱۳۷۵). «چند پیشنهاد دیگر برای اصلاح خط فارسی سرهم‌نویسی و جدانویسی». نگاه نو، ۲۸، ۱۰۱-۱۱۷.
اسلامی، م. (۱۳۸۱). «دشواری‌های پردازش رایانه‌ای خط فارسی». نشر دانش، ۱۹(۳)، ۲۸-۳۲.
اسلامی، م. (۱۳۸۶). «خط فارسی و رسانه‌های گروهی». دوفصلنامه پردازش علائم و داده‌ها، ۸(۲)، ۹۳-۹۸.



- صفا مقدم، الف. (۱۳۸۶). «فاصله‌گذاری در خط فارسی». *نامه فرهنگستان*، ۹ (۴)، ۱۲۳-۱۳۷.
- گل تاجی، م. و برزگر، س. (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». *کتابداری و اطلاع‌رسانی* ۱۳ (۲)، ۱۹۱-۲۱۴.
- محقق‌زاده، م. و زارعیان، ک. (۱۳۸۳). «ارائه راه‌حل برای برخی مسائل اتوماسیون و نگارش فارسی». *فصلنامه اطلاع‌رسانی* ۱۹ (۳ و ۴)، ۱-۱۰.
- معصومی همدانی، ح. (۱۳۸۱). «خط فارسی و رایانه». *نشر دانش*، ۱۹ (۲)، ۲-۶.
- نجفی، الف. (۱۳۸۶). *غلط‌نویسیم: فرهنگ دشواریهای زبان فارسی*. مرکز نشر دانشگاهی، چاپ سیزدهم.
- (1394). "Google Input Tools on Windows".
<https://www.google.com/inputtools/windows/index.html>

