

## طرح کتابخانه رایانه‌ای زبان و ادب فارسی

این کتابخانه اساساً برای استفاده پژوهشگران زبان فارسی، ادبیات، تاریخ، و دیگر رشته‌های علوم انسانی ایجاد خواهد شد که در نهایت، امر پژوهش در زمینه‌های گوناگون فرهنگی، به‌ویژه زبان‌شناسی، دستور زبان فارسی، و فرهنگ بومی را به نحو قابل توجهی تسهیل خواهد کرد و ضریب اطمینان تحلیل‌ها و استنتاج‌ها را فوق‌العاده افزایش خواهد داد. بنابراین جزئیات دستور اجرای این طرح به گونه‌ای خواهد بود که علاوه بر امکان استفاده عموم از آن، قابل استفاده پژوهشگران نیز باشد:

– از آنجا که این طرح تمامی کتاب‌های زبان فارسی (از آغاز تا ۱۳۰۰ هجری شمسی) را شامل است، به جهت زیاد بودن حجم اطلاعاتی که باید وارد نرم‌افزار شوند، ورود اطلاعات در چندین مرحله و به صورت دوره‌ای یا قرن به قرن انجام خواهد شد. مرحله نخست، شامل کتاب‌هایی است که تا نیمه قرن پنجم هجری قمری تألیف شده‌اند. این مجموعه حدوداً پانزده هزار صفحه است.

– انتخاب بهترین چاپ و تصحیح و نیز تهیه شیوه‌نامه حروفچینی به عهده مسئول علمی طرح خواهد بود.

– برای اطمینان بیشتر از درستی و تطابق متن اصلی و متن حروفچینی شده، هر متن را دوبار، دو حروفچین، به طور جداگانه در یک برنامه واژه‌پرداز حروفچینی و ذخیره خواهند کرد. این دو متن حروفچینی شده را یک کارشناس زبان و ادب فارسی، به کمک نرم‌افزاری که برای این کار تهیه خواهد شد، مقابله و با مراجعه به متن اصلی تصحیح خواهد کرد. آنگاه متن تصحیح شده ذخیره، و متن دیگر حذف خواهد شد.

کتابخانه رایانه‌ای زبان و ادب فارسی  
پژوهشگاه زبان و ادب فارسی  
تهران

- در مرحله ورود اطلاعات، در متن هیچ تغییری ایجاد نخواهد شد. حتی غلط‌های چاپی همه تصحیح نمی‌شوند. شماره صفحات نیز مطابق با متن اصلی خواهد بود.
- هر کتاب دارای شناسنامه‌ای خواهد بود که این اطلاعات را دربر خواهد داشت: نام کتاب، نویسنده، مترجم، تاریخ تألیف، محل تألیف، نوع اثر (نظم، نثر، داستانی، علمی و...)، موضوع، نام مصحح، محل چاپ، ناشر، تاریخ چاپ.
- در این طرح، پاورقی‌ها، نسخه‌بدل‌ها، مقدمه مصحح و... هیچ‌کدام حروفچینی و ذخیره نخواهند شد.
- از کتاب ذخیره شده در نرم‌افزار، یک نمونه چاپی تهیه و برای ویرایش و آماده‌سازی متن، به کارشناسان زبان و ادب فارسی سپرده خواهد شد.
- ویرایش و آماده‌سازی متن شامل این موارد است:
۱. تصحیح غلط‌های چاپی و غیرچاپی (به‌ویژه با استفاده از مقالات نقد متون مورد نظر که تا زمان اجرای این طرح در مجلات معتبر به چاپ رسیده‌اند و در اختیار کارشناس قرار خواهند گرفت).
  ۲. تنظیم فاصله‌ها (همه کلمات حتی اجزای کلمات مرکب مانند «زمین خوردن» با یک فاصله حروفچینی می‌شوند).
  ۳. نشانه‌گذاری (علامت سؤال، ویرگول، دو نقطه، ...).
  ۴. تعیین جملات با قرار دادن نقطه در آخر هر جمله (در این برنامه پس از واژه، جمله کوچک‌ترین واحد به شمار می‌آید).
  ۵. اعراب‌گذاری کلمات فارسی، برای ایجاد امکان جست‌وجوی دقیق، و نیز امکان فهرست‌گیری و ایجاد تمایز بین کلمات مشابه در املا و متفاوت در تلفظ و معنا.
  ۶. تعیین عبارات غیرفارسی (عربی شامل آیات، احادیث، سخنان بزرگان، امثال و حکم، شعر و مصراع ترکی؛ لهجه‌ها) که در متون منتخب به کار رفته‌اند.
  ۷. تعیین هریت کلمات و عبارات با کدهای مختلف برای سهولت در فهرست‌گیری‌های زبان لازم شامل فعل، صفت، قید، صوت، نام اشخاص، مکان‌ها، کتاب‌ها، سخنان بزرگان، امثال و حکم، شعر و مصراع.
  ۸. پاراگراف‌بندی و شماره‌گذاری پاراگراف‌های هر کتاب از شماره ۱، برای تسهیل در کار پژوهش و نیز برنامه‌نویسی.
- اعمال موارد فوق در متون ذخیره شده. (متن ذخیره شده اصلی همچنان حفظ

- می‌شود و این اصلاحات در کپی فایل‌های اصلی اعمال می‌شود تا هرگونه تفاوت بین دو متن به آسانی برای کارشناسان طرح، و بعد برای کاربران، قابل تشخیص باشد.
- ۲.۶.۱ - تهیه یک نمونه چاپی پس از اصلاحات یادشده برای بازبینی نهایی و اصلاحات مجدد، که بر عهده مسئول علمی طرح خواهد بود و در صورت لزوم، وی در این مرحله از کارشناسان متخصص زبان فارسی در سطوح کارشناسی ارشد و دکتری ادبیات فارسی دعوت به همکاری خواهد کرد.
- ۲.۶.۲ - اعمال تصحیحات نهایی در متون ذخیره شده.
- تهیه فهرست‌های تفصیلی ساخت‌های پایه و مشتقات آنها. در زبان فارسی حدود چهارصد فعل بسیط داریم که هر کدام تقریباً پنجاه ساخت زمانی دارند؛ و علاوه بر آن، شماره زیادی فعل مرکب هم وجود دارند. فهرست‌هایی باید تهیه شود که ذیل هر مصدر بسیط، ساخت‌های متفاوت آن از قبیل ساخت زمانی، فعل‌های مرکب، و... درج شوند. هر مصدر با زیرمجموعه‌هایش یک بانک کوچک در نرم‌افزار خواهد بود تا جست‌وجوها بر اساس شکل مصدری یا پایه آن به راحتی صورت پذیرد. مثلاً پژوهشگر می‌تواند مشتقات کلمه «انگیختن» را در یک یا چندین متن مورد نظر خود، یا همه متون موجود در نرم‌افزار، جست‌وجو کند، یا می‌تواند با وارد کردن ساخت پایه، فهرست تمام مشتقات آن (مانند برانگیخته بود، انگیزش، برمی‌انگیزم، انگیخته، انگیزه، و...) را با قید شماره صفحه یا پاراگراف و در صورت لزوم جملاتی که این ساخت‌ها در آنها به کار رفته، مشاهده کند. اسم‌ها نیز صورت‌های جمع فارسی و عربی و نیز جمع مکسر دارند که آنها نیز باید ذیل شکل مفرد فهرست شوند. این فهرست‌ها به طور تقریبی بیش از یکصد هزار کلمه خواهند شد که کارشناسان زبان و ادب فارسی در سطوح مختلف کارشناسی، کارشناسی ارشد، و دکتری، باید آنها را تدوین کنند.
- نرم‌افزارهای لازم و ویژگی‌های آنها چنین است:
۱. واژه‌پردازی با فرنت‌های چندگانه فارسی، عربی، دارای نشانه‌های سجاوندی، اعراب و قابلیت مرتب‌سازی حروف و علائم، برای ورود اطلاعات.
  ۲. نرم‌افزار مقابله‌کننده دو متن حروف‌چینی شده.
  ۳. نرم‌افزار اصلی کتابخانه با امکانات زیر:
    - ۳.۱. امکان ذخیره‌سازی اطلاعات متون فارسی با حجم بیش از پانصد هزار صفحه.
    - ۳.۲. قابلیت ایجاد بانک واژگان از همه متون ذخیره شده در نرم‌افزار یا متون انتخابی (که

- کاربر بنا به سلیقه خود انتخاب خواهد کرد) و مرتب کردن آنها به روش های زیر:
- ۱.۳.۲ فهرست کردن الفبایی واژگان بر اساس آغاز واژه همراه با درج شماره بسامد آن در مجموعه مورد جست و جو به دو صورت، یکی از «الف» تا «ی»، و دیگری از «ی» تا «الف».
- ۲.۳.۲ فهرست کردن الفبایی واژگان بر اساس پایان واژه همراه با درج شماره بسامد آن در مجموعه مورد جست و جو به دو صورت، یکی از «الف» تا «ی»، و دیگری از «ی» تا «الف».
- ۳.۳.۲ ارائه فهرست بسامدی واژگان به دو صورت صعودی و نزولی.
- ۴ امکان مشاهده بیت / جمله ای که واژه مورد جست و جو در آن به کار رفته با درج شماره صفحه یا شماره پاراگراف و نام متن یا کد آن.
- ۵ امکان مرتب سازی بیت / جمله های حاوی واژه مورد جست و جو به طرق مختلف و مورد نیاز کاربر، مثلاً مرتب سازی به ترتیب تاریخ تألیف متون.
- ۶ امکان وصل شدن (Link) از «جمله» در فهرست جمله ها به «همان جمله در متن اصلی».
- ۷ امکان انتخاب یک واژه از درون متن اصلی برای مشاهده موارد دیگر کاربرد همان واژه در همان متن و دیگر متون مورد نظر کاربر.
- ۸ امکان جست و جو بر اساس ساخت پایه کلمه (مصدر برای فعل ها، شکل مفرد اسم برای اسم های جمع اعم از جمع فارسی، عربی و مکسر).
- اجرای این بند مستلزم تهیه فهرست های مفصل ساخت های پایه و کلمات مشتقل از آنهاست که توضیح آن پیشتر آمد.
- ۹ امکان شمارش کل کلمات هر متن یا همه متون انتخابی و شمارش کلمات (با حذف بسامد آنها) برای تعیین حوزه واژگان هر متن و در نهایت، واژگان زبان فارسی در هر دوره یا از آغاز تا ۱۳۰۰ هجری شمسی.
- ۱۰ امکان تبدیل متن به جملات متوالی (هر جمله از ابتدای سطر آغاز شود و با نقطه پایان یابد) برای بررسی های زبان شناسی و... و امکان مرتب کردن این جملات / ابیات به صورت الفبایی بنا بر آغاز جمله / بیت و نیز پایان جمله / بیت.
- ۱۱ امکان جست و جوی جملات پرسشی و تعجبی و مرتب کردن تمام جملات پرسشی و تعجبی بر اساس ترتیب الفبایی بنا بر آغاز جمله / بیت و پایان جمله / بیت.
- ۱۲ امکان جست و جوی گفت و گوهای منثور و منظوم و مرتب کردن آنها بر اساس ترتیب تاریخی یا عنوان های انتخاب شده به اختیار کاربر.