

معرفی یک نرم افزار تحلیل حروف ربط، نشانه و اضافه‌ی زبان فارسی

محمد رضا فلاحتی قدیمی فومنی^۱، مرجان حسینی^۲

^۱ دانشیار گروه پژوهشی زبان‌شناسی رایانه‌ای، مرکز منطقه‌ای اطلاع رسانی علوم و فناوری، شیراز، ایران

نویسنده مسئول:

محمد رضا فلاحتی قدیمی فومنی

چکیده

امروزه انجام پردازش‌ها و تحلیل‌های زبانی با استفاده از نرم‌افزار شیوع بسیاری پیدا کرده است. هدف پژوهش حاضر آن بود تا نرم‌افزاری تهیه شود تا بتوان به کمک آن، واژگان زبان فارسی متعلق به شش گروه حروف اضافه‌ی ساده، حروف اضافه مرکب، شبه حروف اضافه، حروف نشانه، حروف ربط و حروف ربط مرکب را شناسایی نمود و برای هر واژه اطلاعاتی آماری شامل فراوانی مطلق، فراوانی نسبی، درصد فراوانی نسبی، فراوانی تجمعی و درصد فراوانی تجمعی را در قالب جداول و نمودارهای خطی و ستونی ارائه کرد. برای انجام کار، فهرست واژگان مربوط به گروه‌های شش‌گانه فوق از کتاب‌های دستور زبان فارسی موجود در بازار استخراج گردید. در مواردی که یک واژه در کتاب‌های مختلف به صورت‌های گوناگونی تحلیل و طبقه‌بندی شده بود، از نظر اکثریت استفاده شد و برای رفع مشکل تنوع املا، برای هر واژه، تمامی صورت‌های املا، پیش‌بینی و فهرست گردید. از معیار فاصله‌ی کامل قبل و بعد از واژه برای تفکیک واژه‌ها (به جز کسره‌ی اضافه) استفاده به عمل آمد و در شناسایی زنجیره‌ی حروف از قاعده‌ی بزرگترین طول استفاده شد. نرم‌افزار طراحی شده با درونداد متون نمونه مورد آزمایش قرار گرفت و خطاهای مشاهده شده، مرتفع گردید. این نرم‌افزار قادر است ضمن تحلیل اطلاعات، نتایج را در قالب فایل اکسل ارائه نماید. این کار، امکان استفاده از جداول و نمودارهای حاصله را به آسان‌ترین شکل در ساختار مقاله، کتاب و آثار دیگر فراهم می‌آورد و در تسریع نگارش مقالات و آثار علمی از سوی محققان حوزه‌ی زبان، ابزاری بسیار مناسب محسوب می‌گردد. دستور نویسان و نحوایان نیز می‌توانند از طریق تحلیل متون خود در نرم‌افزار مربوطه به کاربردی‌تر کردن دستورهای خود بیاورند. نتایج تحلیل متون در نرم‌افزار حاضر از سوی دو متخصص زبان مورد بررسی و بازبینی قرار گرفت و مطابقت نتایج با فهرست‌های واژگان طراحی شده، تأیید شد.

واژگان کلیدی: زبان‌شناسی رایانه‌ای، تحلیل خودکار متن، پردازش زبان فارسی، آموزش زبان، آموزش دستور، نرم‌افزارهای زبانی.

۱- مقدمه

در خصوص حروف اضافه و تحلیل آن کارهای زیادی انجام شده است. مظفری و همکاران (۱۳۹۷) بر اساس قالب‌های معنایی الگوریتمی را ارائه کردند که بتواند از چهار حرف اضافه زبان فارسی (از، در، با و تا) از نظر معنایی ابهام‌زدائی کند. در این پژوهش در مجموع از ۱۰۰۰ جمله داده آموزشی، ۱۰۰ جمله داده توسعه و همچنین ۵۰۰ جمله به عنوان داده تست استفاده شد. در نهایت گزارش شد که الگوریتم تولید شده در ابهام‌زدائی از مفهوم چهار حرف اضافه یاد شده از عملکردی بالای ۹۹ درصد برخوردار بوده است. یه و ویلین (۱۹۹۸) و همچنین باربوسا و همکاران (۲۰۰۷) به بررسی برخی از ویژگی‌های حروف اضافه و ربط در محیط بازیابی اطلاعات پرداختند. القواره و جوسو (۲۰۱۳) نیز در محیط بازیابی اطلاعات به مبحث ابهام‌زدائی از حروف اضافه در ترکیبات حرف اضافه‌ای توجه کردند. اما در پژوهش حاضر هدف آن بود تا نرم‌افزاری تهیه شود تا بتوان به کمک آن، واژگان زبان فارسی متعلق به شش گروه حروف اضافه‌ی ساده، حروف اضافه مرکب، شبه حروف اضافه، حروف نشانه، حروف ربط و حروف ربط مرکب را شناسایی نمود و برای هر واژه اطلاعاتی آماری شامل فراوانی مطلق، فراوانی نسبی، درصد فراوانی نسبی، فراوانی تجمعی و درصد فراوانی تجمعی را در قالب جداول و نمودارهای خطی و ستونی ارائه کرد.

در فارسی برای ایجاد ارتباط بین اجزای جمله و یا جملات مختلف از یک سری واژه‌ها استفاده می‌شود که آنها را حروف یا روابط می‌خوانند. این حروف به دسته‌های مختلفی تقسیم می‌شوند که از آن جمله می‌توان به حروف اضافه، حروف ربط، حرف را و حروف زائد اشاره نمود (شریعت، ۱۳۸۴، ص. ۳۱۱). این گونه حروف یا روابط غالباً دارای معنی مستقلی نیستند و تنها برای ایجاد ارتباط میان اجزای کلام به کار برده می‌شوند. درست به همین خاطر هم هست که گاه برای توصیف این گونه کلمات از الفاظی چون «تهی» استفاده می‌شود. نگاهی گذرا به کتاب‌های دستور زبان فارسی (شریعت، ۱۳۸۴؛ گیوی و انوری، ۱۳۸۸؛ فرشید ورد، ۱۳۸۲؛ لازار، ۱۳۱۷) نشان می‌دهد که دستورنویسان فارسی به حروف نگاهی کاملاً یکدست نداشته‌اند. مثلاً، در یک کتاب، واژه‌ای حرف اضافه‌ی ساده در نظر گرفته شده و در کتابی دیگر حرف اضافه‌ی مرکب، مانند «برای». یا کلمه‌ای در یک کتاب، حرف اضافه است و در کتابی دیگر، حرف ربط و در کتابی دیگر هم حرف اضافه و هم حرف ربط محسوب شده است.

هدف اصلی از پژوهش حاضر ساخت نرم‌افزاری بود تا مقوله‌های شش‌گانه‌ی حروف اضافه‌ی ساده، حروف اضافه‌ی مرکب، شبه حروف اضافه‌ها، حروف ربط ساده، حروف ربط مرکب و حرف نشانه را به صورت خودکار استخراج و تحلیل شود. در ادامه، هر یک از این شش مقوله به اختصار معرفی می‌شود.

۱-۱ حروف اضافه

حروف اضافه کلماتی هستند که قبل از اسم یا به جای اسم می‌آیند و آن را به کلمه‌ی دیگری وابسته می‌کنند. البته، شریعت (۱۳۸۴) اظهار می‌دارد که ممکن است حروف اضافه پس از اسم و یا همزمان در دو موقعیت پیش و پس از اسم نیز ظاهر شوند. متأسفانه، در هیچ یک از کتاب‌های دستوری موجود برای هر مقوله تمامی اعضای موجود در زبان فارسی ذکر نشده و به نمونه‌هایی محدود بسنده شده است. شریعت (۱۳۸۴) حروف اضافه‌ی ساده‌ی زبان فارسی را بدین صورت معرفی می‌کند:

از، الا، اندر، ایدون، به، با، باز، بر، بی، تا، جز، چو، چون، در، را، زی، فا، فرا، فرو، کسره‌ی اضافه، که، مگر، و، وا.

البته، وی اذعان دارد که برخی از حروف اضافه‌ی ساده با حروف ربط و گاهی با قید مشترکند یعنی برحسب عملی که برعهده دارند گاه حرف ربط به شمار می‌روند و گاه قید و گاه حرف اضافه. او به عنوان نمونه‌ای از این دست به الا، ایدون، با، باز، تا، جز، چون، که، مگر و درنهایت و اشاره می‌کند. (ص. ۳۱۲)

۲-۱ حروف اضافه‌ی مرکب

حرف اضافه‌ی مرکب آن است که بیش از یک حرف باشد. حرف‌های وابستگی مرکب بیشتر از بهم‌پیوستگی حرف‌های اضافه و گاهی از بهم پیوستن حرف ربط و اضافه ساخته می‌شوند. شریعت (۱۳۸۴، ص. ۳۱۳) حروف اضافه‌ی مرکب را به طور کلی به دو دسته تقسیم می‌کند:

گروه اول: برای، بجز، بجز از، بی ز، جز از، مگر از.

گروه دوم: بجز که، جز که، چونکه، مگر که، همچون، همچو، همچوکه.

او همچنین اعلام می‌دارد که گاه حرف اضافه‌ی مرکب از پیوستن سه حرف اضافه ساخته می‌شود مانند «از برای» که از «از+ برای+ کسره‌ی اضافه» ساخته شده است. البته، آنچه در کتاب‌های مختلف مشاهده می‌شود نشان می‌دهد که نگاه نویسندگان مختلف به کسره‌ی

اضافه متفاوت است. شریعت آن را واحدی مستقل می‌داند و از این رو از نظر او «برای» یک حرف اضافی مرکب است، حال آنکه برخی دیگر از دست‌نویسان آن را حرف اضافی ساده محسوب می‌دارند و کسره‌ی اضافه را واحدی مستقل فرض نمی‌کنند.

۳-۱ شبه حروف اضافه

شبه حروف اضافه از به هم پیوستن یک یا دو حرف اضافی ساده با یک اسم و گاهی با یک صفت یا یک ضمیر اشاره ساخته می‌شوند. شبه حرف اضافه، عمل یک حرف اضافی ساده را با صراحت و دقت بیشتر انجام می‌دهد. شریعت (۱۳۸۴) در تعیین ملاکی برای شناسایی شبه حرف اضافه بیان می‌دارد که شبه حرف اضافه آن است که بتوان آن را حذف کرد و به جای آن یک حرف اضافی ساده گذاشت. مانند: بهر، از راه، از برای، گذشته از، چنانچون (ص. ۳۱۴). وی همچنین شبه حرف‌های اضافه را به دو گروه تقسیم می‌کند: گروه اول از به هم پیوستن یک اسم یا یک صفت یا یک ضمیر با یک حرف اضافه ساخته می‌شود و مهم‌ترین آنها عبارتند از: بر، بهر، بیرون، پی، پیش، جهت، دون، سوی، غیر، گذشت، مانند، مانده، مثل، نزدیک (همه به کسر آخر) بیرون ز، جدا از، چنان، غیر از، گذشت از، گذشته از.

گروه دوم، بیشتر از بهم پیوستن یک اسم با دو حرف اضافه ساخته می‌شود بدین ترتیب که حرف نخستین را پیش از اسم و حرف دوم را که معمولاً کسره‌ی اضافه است پس از آن می‌آورند. شبه حرف‌های مهم این گروه از این قرار است: از بهر، از پی، از جهت، از راه، از روی، از سر، از قبل، از واسطه، اندر جنب، باب، به جای، به جهت، بدون، برای، بسان، به سر، به سوی، بغیر، به کردار، به نزدیک، باسر، برجای، برسان، بر سر، بر کردار، درباب، در جنب، در حق (هه به کسر آخر) و چنانچون (ص. ۳۱۵).

۴-۱ حروف ربط

از نظر شریعت (۱۳۸۴) حرف ربط کلمه‌ای است که دو کلمه یا دو جمله را به هم می‌پیوندد مانند: و، یا، تا، که... در جمله‌ی «تقی و اکبر آمدند» حرف ربط «و» دو کلمه را به هم ربط داده است. اما حرف ربط می‌تواند دو جمله را نیز به هم پیوند دهد که در این صورت سه حالت ممکن است اتفاق بیافتد:

الف- دو جمله مستقل هستند و حرف ربط آن دو جمله مستقل را به هم می‌پیوندد مانند: «حسن رفت و تقی آمد» (شریعت، ۱۳۸۴، ص. ۳۲۰). مهم‌ترین حروف این دسته عبارتند از: و، یا، ولی، اما، بلکه، تا، پس، سپس، با، هم، نیز، لیکن، لکن، لیک، با این همه، همچنین، حتی، وانگهی، مع هذا، با این حال، در نتیجه، بنابراین، وگرنه، چون، برعکس، والا، چه... چه... خواه... خواه و ...
ب- جمله‌ای که با حرف ربط درست شده است یکی از ارکان جمله‌ی دیگر می‌شود مانند: «من دیدم که او از خانه بیرون آمد». در این مورد عبارت «او از خانه بیرون آمد» مفعول بی واسطه‌ی جمله‌ی قبلی یعنی «من دیدم» است که به وسیله‌ی حرف ربط «که» به جمله‌ی قبلی مربوط شده است. مهم‌ترین حروف این دسته عبارتند از: که، چون، کجا، وقتی، چنانکه، تا، به طوری که، تا اینکه، اگر، اگرچه، هرگاه، در صورتی که، مگر این که، الا این که، زیرا، به علت این که، هرچه، هر قدر، هر کجا، چندان که، هر چند و...
ج- جمله‌ای که با حرف ربط شروع شده است، کلمه‌ای از جمله‌ی دیگر را تشریح می‌کند، مانند: «آن پسری که دیدی برادر من بود». در این عبارت «آن پسر برادر من بود» جمله‌ی اصلی است و لفظ «دیدنی» که با حرف «که» آغاز شده است، کلمه‌ی پسر را تشریح می‌کند. مهم‌ترین حروف ربط دسته سوم عبارتند از: که، کجا، تا (شریعت، ۱۳۸۴، ص. ۳۲۱).

۵-۱ حروف ربط مرکب

حرف ربط را از نظر ساختمان نیز به دسته‌های مختلفی تقسیم می‌کنند که عبارتند از:
الف- حروف ربط ساده که پیشتر معرفی شد فقط یک جزء دارند مانند: و، اگر، یا، تا.
ب- حروف ربط مرکب که خود به دو گروه تقسیم می‌شوند، گروهی که از دو جزء ساخته شده باشند و دو جزء استقلال خود را از دست داده باشند. مانند: همینکه، بلکه و گروهی که از دو یا چند کلمه بدون از دست دادن استقلال آنها درست شده باشند و به آنها می‌توان گروه ربطی گفت، مانند: به منظور این که، به علت این که، وقتی که (ص. ۳۲۳).

۶-۱ حروف نشانه

نشانه‌ها در فارسی متعدّدند اما برای پرهیز از ابهام و امکان تحلیل آماری بدون تحلیل‌های معناشناختی پیچیده صرفاً حرف نشانه‌ی «را» در پژوهش حاضر مورد بررسی قرار گرفت.

۲- الگوریتم برنامه

در این بخش، زیر ساخت نرم افزار به ترتیب مراحل اجرایی تشریح می گردد:

۲-۱ دریافت متن

نرم افزار حاضر به گونه ای طراحی گردیده است که می تواند فایل های دارای پسوند doc، docx و txt را دریافت نماید. البته، در نرم افزار جعبه ای طراحی شده که می توان بخشی از فایل متنی را کپی و در آن وارد نمود.

۲-۲ تبدیل فایل به فرمت txt.

چنانچه فایل ورودی به فرمت txt نباشد (به فرمت doc یا docx باشد) در این مرحله به عنوان نوعی پیش پردازش، تمامی اطلاعات متن جز واژگان از ساختار فایل حذف می گردد. بدین ترتیب، تمامی نمودارها و اطلاعات فرامتنی از متن کنار گذاشته می شود و تنها متن مربوطه باقی می ماند. اطلاعات باقی مانده، با پسوند txt ذخیره می شود.

۲-۳ فهرست کلمات

با توجه به اینکه در نرم افزار حاضر شش گروه کلمات شامل حروف اضافه ی ساده، حروف اضافه ی مرکب، شبه حروف اضافه، حروف ربط ساده، حروف ربط مرکب و حروف نشانه، مورد پردازش قرار می گیرند، فهرستی از اعضای هر گروه توسط محقق و با استفاده از کتاب های دستور زبان فارسی تهیه شد. این اطلاعات به عنوان «فهرست کلمات» در قالب شش فهرست جدا و یک فهرست درهم کرد تهیه شد و در ساختار نرم افزار قرار گرفت. با توجه به وجود انواع مختلف فاصله ها و شیوه های پیوند واژه در فارسی، تمامی صورت های احتمالی هر واژه توسط محقق پیش بینی و فهرست شد. مزیت این کار آن است که می توان تمامی رخداد های یک واژه را (با هر صورت ظاهری که نوشته شده باشد) شناسایی، فهرست و درهم کرد نمود. البته، اگرچه صورت های مختلف شناسایی می شود، تنها از یک صورت ظاهری به عنوان نشان دهنده ی مدخل که حاوی درهم کرد اطلاعات آماری اشکال مختلف آن واژه است، در بخش تحلیل اطلاعات استفاده می شود. بدین ترتیب، «از این رو» و «از اینرو» هر دو مورد شناسایی قرار می گیرند. منتهی برای نمایش نتیجه ی نهایی تنها از یکی از این دو صورت استفاده می شود و آماری که برای این صورت منتخب ارایه می شود در بردارنده ی آماری است که برای تمامی صورت های ظاهری مختلف در متن بدست آمده است.

۲-۴ تحلیل متن

تمامی واژه های موجود در فهرست کلمات در فایل متنی جستجو می شوند و برای هر واژه پنج متغیر محاسبه می شود که عبارتند از: (الف) فراوانی مطلق، (ب) فراوانی نسبی، (ج) درصد فراوانی نسبی، (د) فراوانی تجمعی و (ه) درصد فراوانی تجمعی. شیوه ی محاسبه ی متغیرهای فوق به شرح زیر است:

الف- فراوانی مطلق: این مختصه تعداد دفعاتی را نشان می دهد که یک واژه در متن مربوطه ظاهر شده است.

ب- فراوانی نسبی: این مختصه از طریق تقسیم فراوانی مطلق بر تعداد کل واژگان موجود در متن محاسبه می شود. به عنوان مثال، اگر واژه ی «به» ۱۰ بار در متنی که حاوی ۱۰۰ واژه بوده ظاهر شود، فراوانی نسبی این واژه در متن برابر با ۱/۱۰ خواهد بود.

ج- درصد فراوانی نسبی: از ضرب عدد بدست آمده برای فراوانی نسبی در ۱۰۰ بدست می آید. مثلاً چنانچه فراوانی نسبی بدست آمده یک واژه ۱/۱۰ باشد، درصد فراوانی نسبی آن برابر با $100 \times 1/10 = 10$ خواهد بود.

د- فراوانی تجمعی: عدد این مختصه در پله ی اول همان عدد فراوانی مطلق است. برای محاسبه ی فراوانی تجمعی در پله ی دوم فراوانی مطلق در پله ی اول و دوم با هم جمع می شود. در پله ی سوم، فراوانی تجمعی از طریق افزودن فراوانی مطلق در پله های یک، دو و سه بدست می آید و به همین شکل تا آخر. فراوانی تجمعی میزان حجمی را که واژگانی خاص در یک متن به خود اختصاص می دهند نشان می دهد.

ه- درصد فراوانی تجمعی: این مختصه در پله ی اول از طریق ضرب فراوانی نسبی پله ی مربوطه در ۱۰۰ بدست می آید. مثلاً اگر حرف «و» فراوان ترین واژه و دارای فراوانی نسبی ۰/۰۴۵۳ بوده باشد، درصد فراوانی تجمعی در همین پله معادل ۴/۵۳ خواهد بود. حال اگر واژه ی ردیف دو «به» و فراوانی نسبی آن ۰/۰۲۳۷ باشد درصد فراوانی تجمعی برای این دو پله از طریق ضرب فراوانی نسبی در هر پله در ۱۰۰ و سپس جمع نتایج محاسبه می شود.

پله ۱: «و» فراوانی نسبی	۰/۰۴۵۳	درصد فراوانی تجمعی	۴/۵۳
پله ۲: «به» فراوانی نسبی	۰/۰۲۳۷	درصد فراوانی تجمعی	۲/۳۷ + ۴/۵۳
پله ۳: «در» فراوانی نسبی	۰/۰۲۳۷	درصد فراوانی تجمعی	۲/۳۷ + ۴/۵۳ + ۲/۳۷

۲-۵ تعیین نوع مرتب‌سازی خروجی

پیش از تحلیل اطلاعات کاربر می‌تواند نوع مرتب‌سازی خروجی را تعیین نماید. در نرم‌افزار حاضر، سه نوع مرتب‌سازی وجود دارد که کاربر براساس نیاز خاص خود می‌تواند یکی را آن‌ها را به شرح زیر برگزیند: (الف) براساس حروف الفبا، (ب) براساس فراوانی مطلق به صورت صعودی و (ج) براساس فراوانی مطلق بصورت نزولی.

۲-۶ اجزای خروجی و نحوه نمایش

اجزای خروجی از راست به چپ ردیف، واژه، فراوانی مطلق، فراوانی نسبی، درصد فراوانی نسبی، فراوانی تجمعی و درصد فراوانی تجمعی را شامل می‌گردد.

نحوه نمایش خروجی نیز به دو شیوه می‌باشد. در شیوه اول تمامی اطلاعات مربوط به گروه‌های شش‌گانه (حروف اضافه‌ی ساده، حروف اضافه‌ی مرکب، شبه حروف اضافه، حروف نشانه، حروف ربط ساده و حروف ربط مرکب) به صورت درهم‌کرد و در قالب جدولی واحد ارایه می‌شود. در شیوه دوم اطلاعات مربوط به هر یک از گروه‌های شش‌گانه در قالب جدولی مجزا ارایه می‌گردد.

۲-۷ تولید جدول و نمودار در قالب فایل اکسل

نحوه طراحی نرم‌افزار به گونه‌ای است که می‌تواند بنا به درخواست کاربر، اطلاعات بدست آمده را در قالب جداول و نمودارهایی ارایه نماید. این گونه اطلاعات در قالب فایل اکسل ارایه می‌شود و کاربر می‌تواند از آنها روگرفت تهیه نموده در متن، مقاله و یا اثر خود کپی نماید.

۲-۸ چند نکته‌ی دیگر

- به عنوان بخشی از عملیات پیش‌پردازش، از «فاصله کامل» (full space) به عنوان معیاری برای تفکیک کلمات در متن اصلی استفاده شده است. با این همه، چون در زبان فارسی کلمات علاوه بر فاصله کامل با علائم نگارشی و نشانه‌گذاری نیز از هم تفکیک می‌شوند، در مرحله‌ی پیش‌پردازش، علائم نشانه‌گذاری زیر با فاصله‌ی کامل جایگزین شد:

علائم نشانه‌گذاری: ، - * + / ؟ ! ؛ : .	
اعداد: 1 2 3 4 5 6 7 8 9 0	
\n	خط جدید
\t	Tab
\r	فاصله بزرگ

- همانگونه که پیشتر نیز ذکر شد برای جداسازی یک واژه، از وجود فاصله‌ی کامل قبل و بعد از واژه استفاده گردید. اما کسره‌ی اضافه دارای ویژگی خاصی است، بدین ترتیب که به کلمه‌ی پیشین خود اتصال دارد. به همین منظور، برای شناسایی کسره‌ی اضافه (-□)، عدم وجود فاصله قبل از واژه و وجود فاصله‌ی کامل پس از آن به عنوان معیار مورد استفاده قرار گرفت. بدین ترتیب، نرم‌افزار توانست واژه‌ی «برای» را ترکیبی از «برای + کسره‌ی اضافه» شناسایی نماید.

- از آنجایی که در خط فارسی در بسیاری از موارد، امکان نگارش یک واژه به بیش از یک شکل وجود دارد (مانند، «به جای» در مقابل «جای»، «هم چنین» در مقابل «همچنین»)، نرم‌افزار به گونه‌ای طراحی شد که بتواند این گونه‌های صوری مرتبط را در قالب مدخلی واحد تحلیل و گزارش نماید. تنوع نگارش ناشی از اتصال و انفصال و نیز به‌کارگیری انواع مختلف فاصله و یا نشانه‌های زبرنجیری از سوی محقق برای هر واژه فهرست شد و در واقع برای هر واژه تمامی حالت‌های مختلف احتمالی که واژه می‌توانست با آن در متن ظاهر شود مشخص شد. در گام بعدی، نرم‌افزار پس از شناسایی هر صورت و ارایه‌ی اطلاعات آماری آن، پیش از ارایه‌ی تحلیل نهایی

تمامی اطلاعات آماری صورت‌های مرتبط را در هم کرد نموده و ذیل یک صورت معیار نشان می‌دهد. برای مثال، اگر «هم چنین» با فراوانی ۴ و «همچنین» با فراوانی ۵ در متنی ظاهر شده باشد، نرم‌افزار یکی از این دو صورت را به عنوان واژه‌ی مدخل برمی‌گزیند و عدد ۹ را مقابل آن یادداشت می‌کند.

- با توجه به اینکه در گروه‌های شش‌گانه مورد بررسی، حروف بسیط و مرکب وجود داشت، لازم بود در خصوص تشخیص این گونه واژه‌ها در متن سیاست‌گذاری شود. به عبارت دیگر، آیا «با این که» می‌بایست یکبار و فقط یکبار به عنوان حرف ربط مرکب در نظر گرفته می‌شد یا علاوه بر این می‌بایست اجزای آن نیز در گروه‌های خاص خود قرار می‌گرفتند. بدون سیاست‌گذاری و با توجه به معیارهایی که برای تفکیک واژه‌ها در نظر گرفته شده بود، امکان پردازش مکرر یک واژه در متن وجود داشت که در صورت بروز چنین اتفاقی، آمار ارایه شده نمی‌توانست واقعی و قابل اعتماد باشد. از این رو، برای رفع این مشکل قاعده‌ای تعریف گردید بدین صورت که معیار اصلی در تفکیک واژه‌ها، معیار «بیشترین طول» باشد. به بیان ساده‌تر، در تحلیل عبارت «از برای» نرم‌افزار ابتدا «از» را تشخیص می‌دهد. بلافاصله «برای» نیز که پیش و پس از آن فاصله‌ی کامل آمده، شناسایی می‌شود. با استفاده از قاعده‌ی تعریف شده، نرم‌افزار به جای اینکه «از» و «برای» را دو مدخل فرض کند، به دلیل اینکه واژه‌های دیگر بین آن دو نیامده، و چون هر دو واژه در فهرست گروه‌های شش‌گانه وجود دارد، آنها را یک ترکیب فرض می‌کند و ترکیب شناسایی شده را یکبار و فقط یکبار تحلیل می‌کند. بنابراین، شرط اصلی برای شناسایی همواره رویت بزرگترین طول عبارت خواهد بود.
- سرعت پردازش اطلاعات در نرم‌افزار تا حد زیادی به RAM و قدرت رایانه‌ی شخصی بستگی دارد. در یک رایانه‌ی معمولی با RAM معادل 2GB این نرم‌افزار می‌تواند مراحل پردازش متنی با طول ۱۰۰ صفحه را در "۴۰:۱" به انجام برساند. متنی با ۳۰۰ صفحه زمانی معادل "۴۰:۴" نیاز خواهد داشت.

۹-۲ شناسایی و رفع خطاهای برنامه

در مراحل ساخت و ارزیابی نرم‌افزار خطاهایی به شرح زیر مشاهده و مرتفع گردید:

- مشاهده شد که نمایش نتایج در قالب نمودار براساس تمامی مختصات بسیار پیچیده است و به نمودارهایی بسیار طولانی نیاز دارد که در صورت حفظ خوانا بودن نمودارها، در یک صفحه عادی نمی‌گنجد. از این رو، دو تدبیر اندیشیده شد. (۱) نمودارها صرفاً براساس متغیر «فراوانی مطلق» تهیه شدند و (۲) تمامی واژگان دارای فراوانی صفر از نمودارها حذف شدند.
- از آنجا که واژگان گروه‌های شش‌گانه از کتاب‌های دستوری مختلف استخراج گردیده بود و در این کتاب‌ها (و حتی گاه در یک کتاب) برخی موارد به عنوان اعضای گروه‌های مختلف معرفی گردیده بودند، با مشورت یک متخصص زبان فارسی برای پرهیز از ابهام و جلوگیری از عضویت یک واژه در چند گروه، عضویت این دسته از واژگان در گروه‌های متعدد حذف و هر واژه صرفاً در یک گروه قرار گرفت.
- بخشی تحت عنوان «ارسال پیشنهاد» به نرم‌افزار اضافه گردید.
- به خاطر استفاده‌ی از معیار فاصله پیش و پس از واژه جهت تقطیع متن، نرم‌افزار اولیه قادر به جداسازی کسره اضافه (-) نبود. از این رو، قاعده‌ای تعریف شد تا کسره‌ی اضافه نیز استخراج شود. در این قاعده، پیش از کسره‌ی اضافه فاصله‌ای وجود ندارد و فقط لازم است پس از آن یک فاصله‌ی کامل وجود داشته باشد.

۱۰-۲ ارزیابی برنامه

دو متخصص زبان با مدرک کارشناسی‌ارشد زبان انگلیسی از طریق درونداد متونی کوتاه به برنامه نتایج را با فهرست‌های شش‌گانه‌ی تولید شده مطابقت دادند و صحت عملکرد نرم‌افزار را از طریق مقایسه‌ی تحلیل به صورت دستی با تحلیل خودکار تایید نمودند.

۳- توصیف گام به گام نرم افزار



تصویر ۱: نمایش صفحه اصلی نرم افزار.

تصویر ۱ نمایشی کلی از صفحه اصلی سامانه را نشان می دهد. این صفحه سه بخش عمده‌ی صفحه اصلی، تحلیل متن و ارسال پیشنهاد را شامل می شود. در صفحه‌ی اصلی، اطلاعاتی کلی در خصوص نرم افزار و پیشینه‌ی طرح ارائه گردیده است.



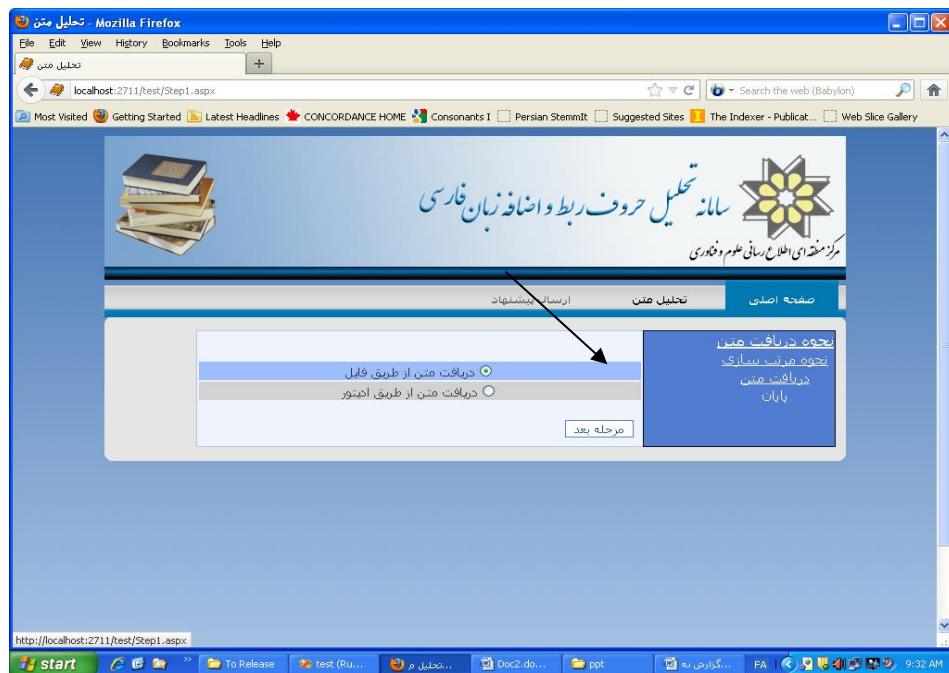
تصویر ۲: نمایش جعبه‌ی ارسال پیشنهاد برای سامانه.

تصویر ۲ بخش ارسال پیشنهادات نرم افزار را نشان می دهد. کاربران پس از درونداد و تحلیل متن خود و ارزیابی آن می توانند نظرات و پیشنهادات سازنده‌ی خود را برای مجری ارسال دارند. ضمناً، در صورت وجود هرنوع ابهام یا پرسش و نیاز به راهنمایی می توانند پیام مربوطه را از این طریق ارسال دارند. در این بخش، اطلاعات مربوط به فرد تماس گیرنده (شامل نام و نام خانوادگی و آدرس پست الکترونیکی) در سیستم ثبت می شود و بسته به مورد پاسخ مربوطه برای کاربر ارسال می شود.



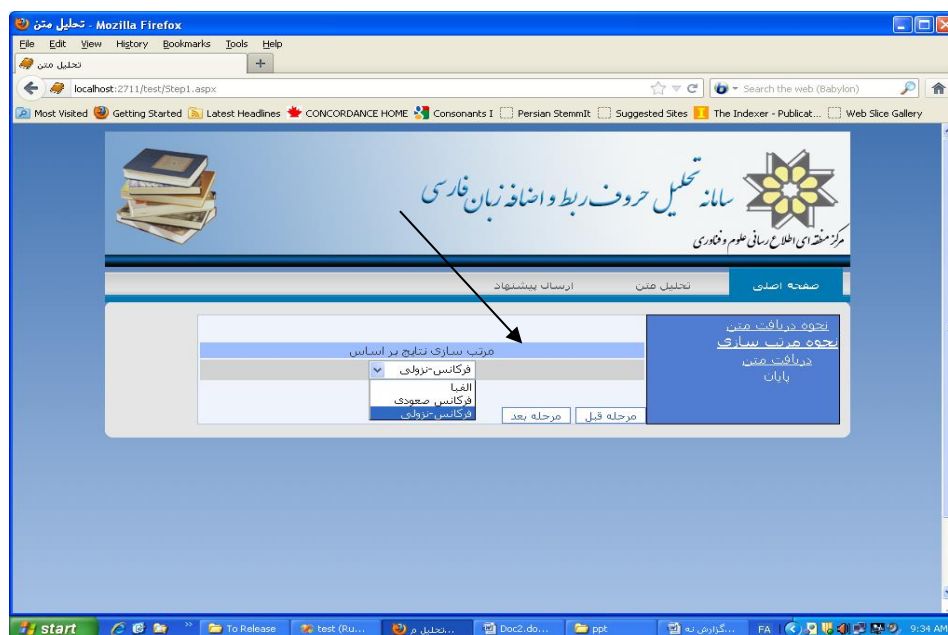
تصویر ۳: نمایش مرحله‌ی اول تحلیل متن در نرم‌افزار.

تصویر ۳ گزینه‌ی تحلیل متن را در نرم‌افزار نشان می‌دهد. برای آغاز تحلیل متن لازم است نخست روی گزینه‌ی تحلیل متن کلیک زده شود.



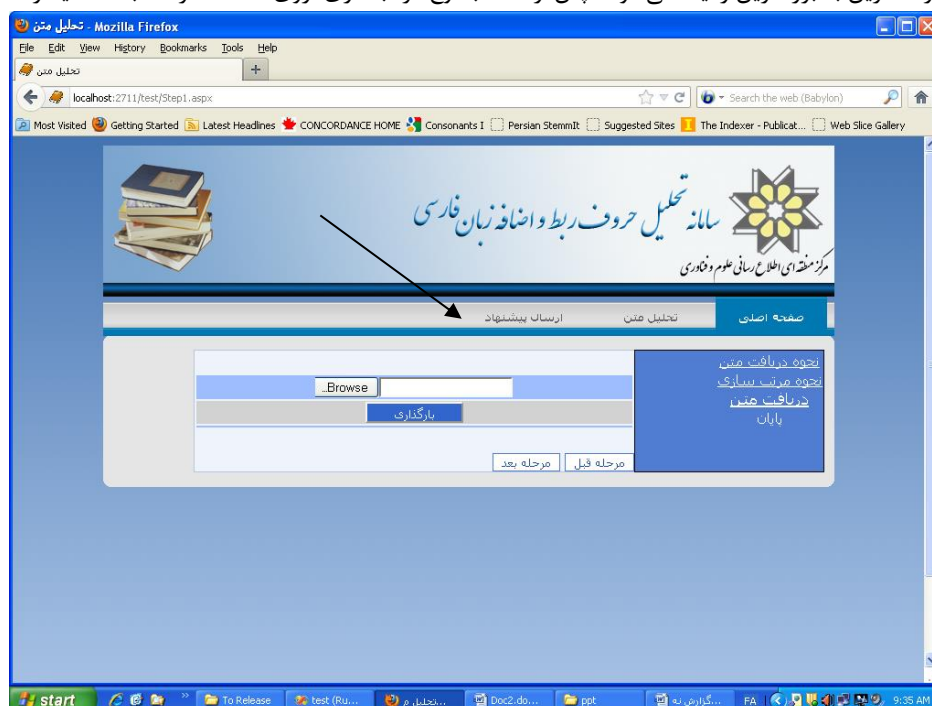
تصویر ۴: نمایش مرحله‌ی دوم تحلیل متن در نرم‌افزار.

پس از کلیک زدن بر روی گزینه‌ی تحلیل متن در تصویر ۳، اطلاعاتی که در تصویر ۴ نشان داده شده است، رویت خواهد گردید. در این مرحله دو امکان وجود دارد که می‌توان یکی از آنها را برگزید. با انتخاب گزینه‌ی «دریافت متن از طریق فایل» کاربر می‌تواند فایلی را که در رایانه‌ی شخصی خود به صورت یک فایل doc یا docx یا txt ذخیره کرده انتخاب نماید. در صورتی که کاربر گزینه‌ی «دریافت متن از طریق ادیتور» را برگزید می‌تواند بخش یا تمام اطلاعات موجود در یک فایل دارای پسوند doc یا docx یا txt را کپی و در جعبه‌ی ادیتور وارد نماید. پس از انتخاب یکی از دو حالت فوق بر روی دکمه‌ی «مرحله بعد» کلیک زده می‌شود.



تصویر ۵: نمایش مرحله‌ی دوم تحلیل متن در نرم‌افزار.

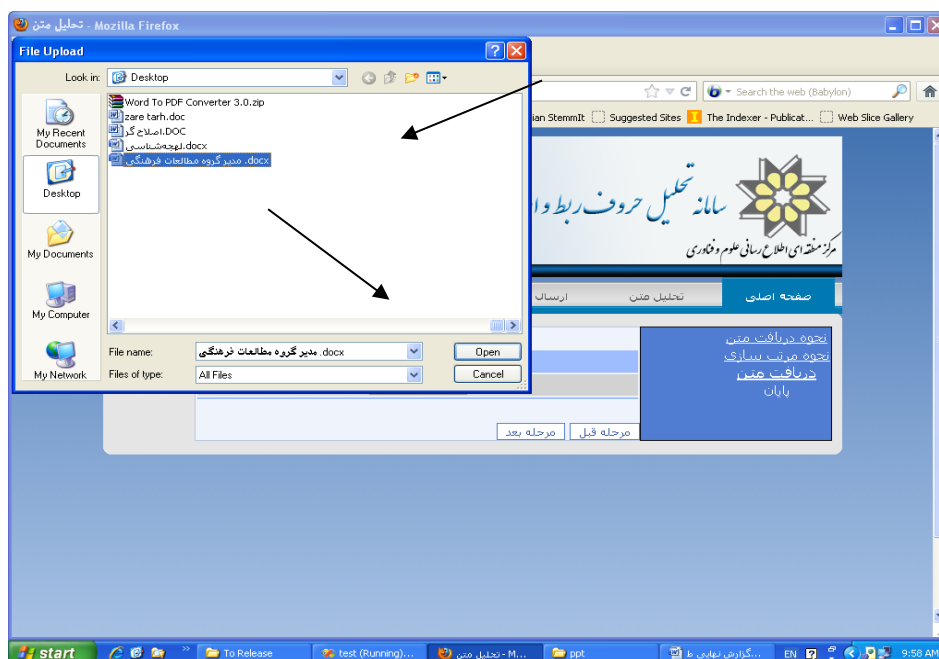
پس از انتخاب متن مورد پردازش (انتخاب فایل و یا کپی بخشی از فایل و ثبت آن در جعبه‌ی ادیتور)، لازم است نحوه‌ی مرتب‌سازی نتایج مشخص گردد. برای مرتب‌سازی نتایج سه انتخاب وجود دارد که عبارتند از: الفبایی، فرکانس نزولی و فرکانس صعودی. در مرتب‌سازی الفبایی نتایج بدون توجه به فراوانی رخداد و صرفاً براساس ترتیب الفبایی ردیف می‌شوند. در فرکانس نزولی واژه‌ها براساس فراوانی رخداد از فراوان‌ترین به نادرترین واژه ردیف می‌شوند. فرکانس صعودی کاملاً عکس حالت فوق است، بدین معنی که مدخل‌ها براساس فراوانی از کمترین به بزرگترین ردیف می‌شوند. پس از انتخاب نوع مرتب‌سازی، روی دکمه «مرحله بعد» کلیک زده می‌شود.



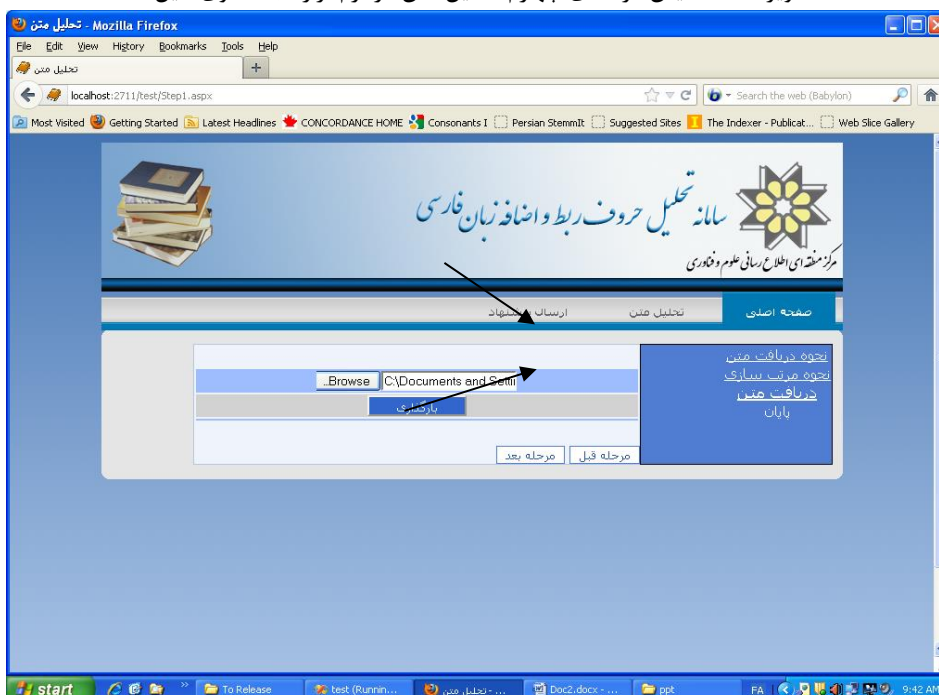
تصویر ۶: نمایش مرحله‌ی سوم تحلیل متن در نرم‌افزار.

پس از انتخاب شیوه‌ی مرتب‌سازی لازم است فایل مربوطه انتخاب و بارگذاری شود. با توجه به اینکه در مراحل پیشین گزینه‌ی «دریافت متن از طریق فایل» انتخاب شده بود، حال می‌توان فایلی را که دارای پسوند doc یا docx و یا txt باشد از رایانه‌ی شخصی

انتخاب نمود. برای این کار روی گزینه‌ی "Browse" کلیک زده می‌شود و از پوشه مربوطه و موجود در رایانه‌ی شخصی فایل مورد نظر انتخاب می‌شود.

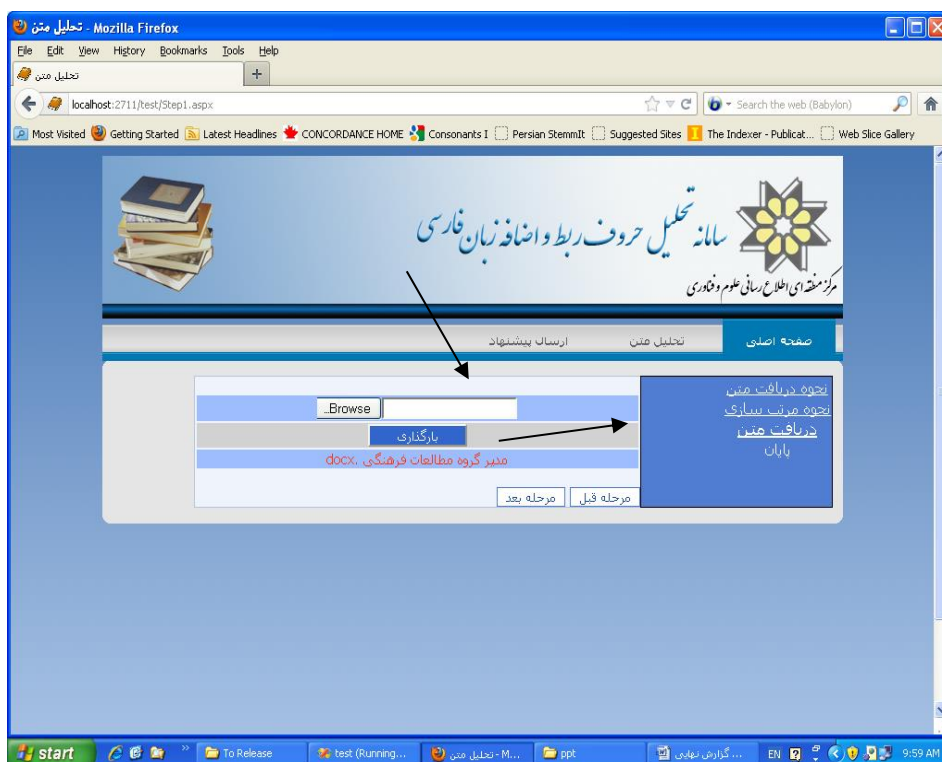


تصویر ۷-۱. نمایش مرحله‌ی چهارم تحلیل متن در نرم‌افزار (جستجوی فایل).



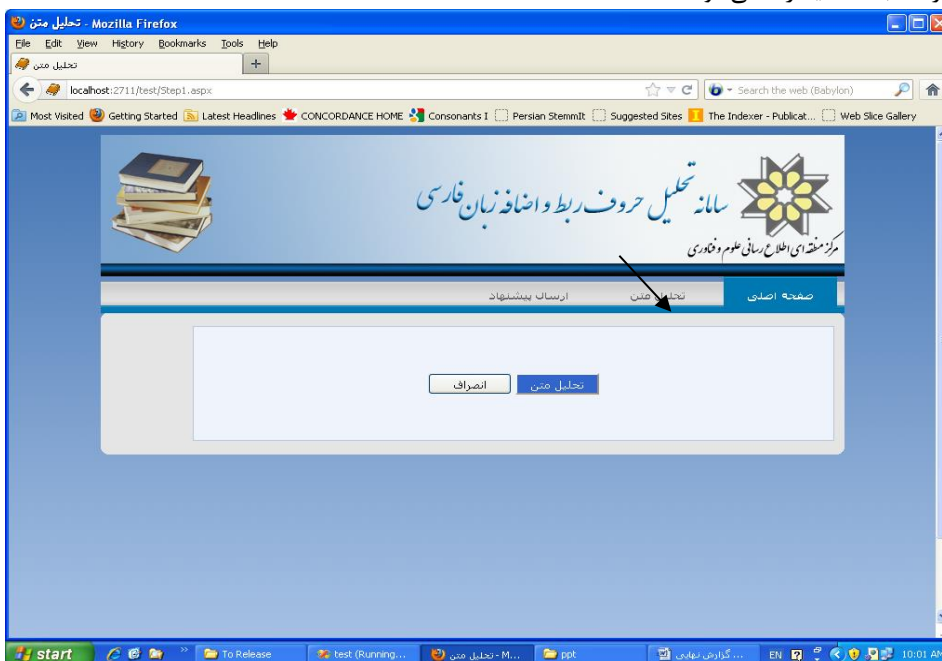
تصویر ۷-۲. نمایش مرحله‌ی چهارم تحلیل متن در نرم‌افزار (انتخاب فایل).

پس از انتخاب فایل از پوشه مربوطه، مسیر فایل بارگذاری شده در جعبه‌ی روبروی گزینه‌ی "Browse" نشان داده می‌شود. سپس، بر روی دکمه‌ی «بارگذاری» کلیک زده می‌شود.



تصویر ۸: نمایش مرحله ی پنجم تحلیل متن در نرم افزار.

پس از کلیک زدن بر روی دکمه ی «بارگذاری» در تصویر ۷، فایل بارگذاری شده به صورتی که در تصویر ۸ مشاهده می شود، نمایش داده می شود. به عنوان مثال، فایل بارگذاری شده در تصویر ۸، فایل «مدیر گروه مطالعات فرهنگی. docx» است. پس از انجام بارگذاری، روی دکمه ی «مرحله بعد» کلیک زده می شود.



تصویر ۹: نمایش مرحله ی ششم تحلیل متن در نرم افزار.

در این مرحله، نرم افزار آماده ی انجام تحلیل متن است. برای این منظور تنها کافی است بر روی دکمه ی «تحلیل متن» کلیک زده شود. چنانچه کاربر از ادامه ی کار منصرف شده باشد، می تواند دکمه ی «انصراف» را انتخاب نموده، از برنامه خارج شود.

ردیف	واژه	فراوانی مطلق	فراوانی نسبی	درصد فراوانی نسبی	درصد فراوانی تجمعی	درصد فراوانی تجمعی
1	و	399	0.0644	6.4417	399	6.4417
2	در	199	0.0321	3.2128	598	9.6545
3	به	180	0.0291	2.9060	778	12.5605
4	از	165	0.0266	2.6639	943	15.2244
5	را	125	0.0202	2.0181	1068	17.2425
6	که	93	0.0150	1.5015	1161	18.7440
7	با	74	0.0119	1.1947	1235	19.9387
8	بر	25	0.0040	0.4036	1260	20.3423
9	تا	19	0.0031	0.3067	1279	20.6490
10	پس از	15	0.0024	0.2422	1294	20.8912
11	اما	14	0.0023	0.2260	1308	21.1172
12	نیز	14	0.0023	0.2260	1322	21.3432

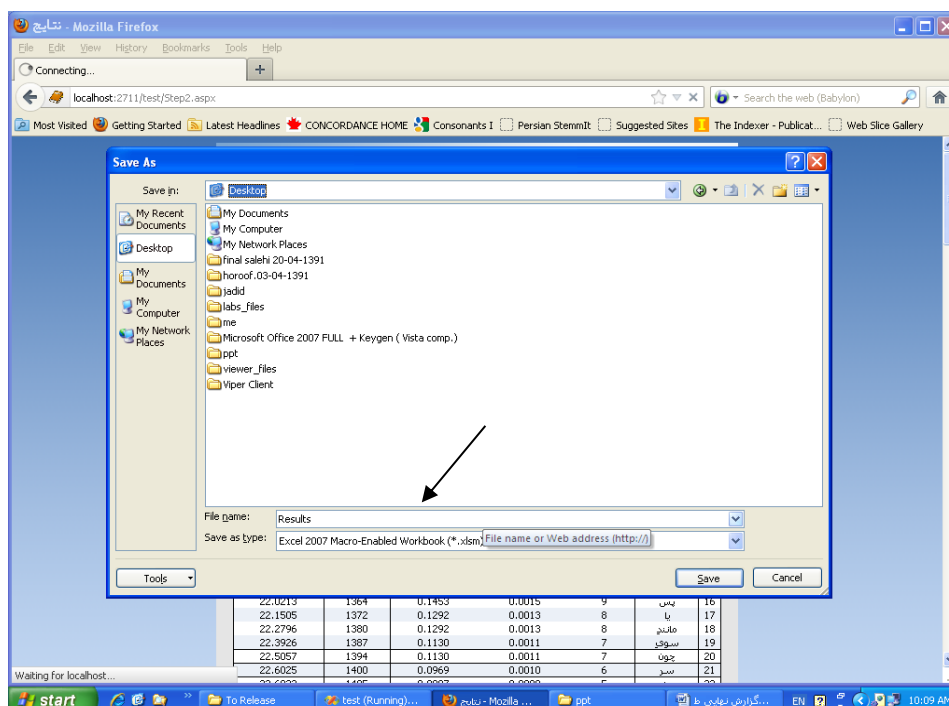
تصویر ۱-۱۰: نمایش مرحله‌ی هفتم تحلیل متن در نرم‌افزار (نمایش نتایج به صورت کلی).

پس از انتخاب دکمه‌ی «تحلیل متن» در نمودار ۹، جدولی حاوی نتایج تحلیل تولید می‌گردد. نحوه‌ی نمایش نتایج در این نرم‌افزار به دو صورت است: ۱- به صورت کلی (تصویر ۱-۱۰) و ۲- به صورت مجزا (تصویر ۱-۱۰). در واقع، می‌توان اطلاعات آماری مربوط به واژگان موجود در گروه‌های شش‌گانه را به صورت جدا محاسبه نمود و یا همه‌ی آنها را در قالب یک جدول و فهرست واحد، نشان داد. پیش فرض در نرم‌افزار، «نمایش نتایج به صورت مجزا» تعیین گردیده که البته کاربر در صورت تمایل می‌تواند «نمایش نتایج به صورت کلی» را برگزیند. نرم‌افزار حاضر به گونه‌ای طراحی گردیده است که می‌تواند از اطلاعات آماری تولید شده، خروجی اکسل نیز تولید نماید. امکان دریافت اطلاعات در قالب جدول و نمودار نیز وجود دارد. برای دریافت اطلاعات در قالب فایل اکسل باید بر روی گزینه‌ی «دریافت فایل اکسل» کلیک زد.

ردیف	واژه	فراوانی مطلق	فراوانی نسبی	درصد فراوانی نسبی	درصد فراوانی تجمعی	درصد فراوانی تجمعی
1	هم چون	1	0.0002	0.0161	1471	23.7488
2	هم جو	0	0.0000	0.0000	1472	23.7649
3	مگر از	0	0.0000	0.0000	1472	23.7649
4	جز که	0	0.0000	0.0000	1472	23.7649
5	جز از	0	0.0000	0.0000	1472	23.7649
6	به جز	0	0.0000	0.0000	1472	23.7649
7	چون که	0	0.0000	0.0000	1472	23.7649
8	به جز از	0	0.0000	0.0000	1472	23.7649
9	هم چون که	0	0.0000	0.0000	1472	23.7649
10	به جز که	0	0.0000	0.0000	1472	23.7649
11	مگر که	0	0.0000	0.0000	1472	23.7649
12	از پس از	0	0.0000	0.0000	1472	23.7649
	مجموع	1	0.0002 = (1/6194)	0.0161	1472	23.7649

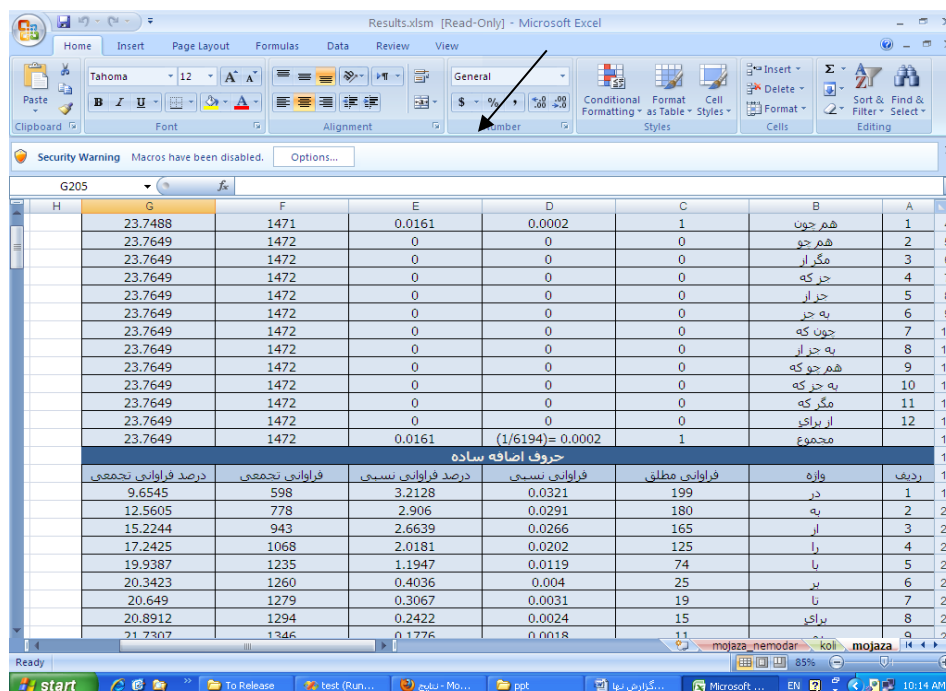
تصویر ۲-۱۰: نمایش مرحله‌ی هفتم تحلیل متن در نرم‌افزار (نمایش نتایج به صورت مجزا).

در نرم‌افزار حاضر، برای هر واژه اطلاعات زیر نمایش داده می‌شود: فراوانی مطلق، فراوانی نسبی، درصد فراوانی نسبی، فراوانی تجمعی و درصد فراوانی تجمعی.



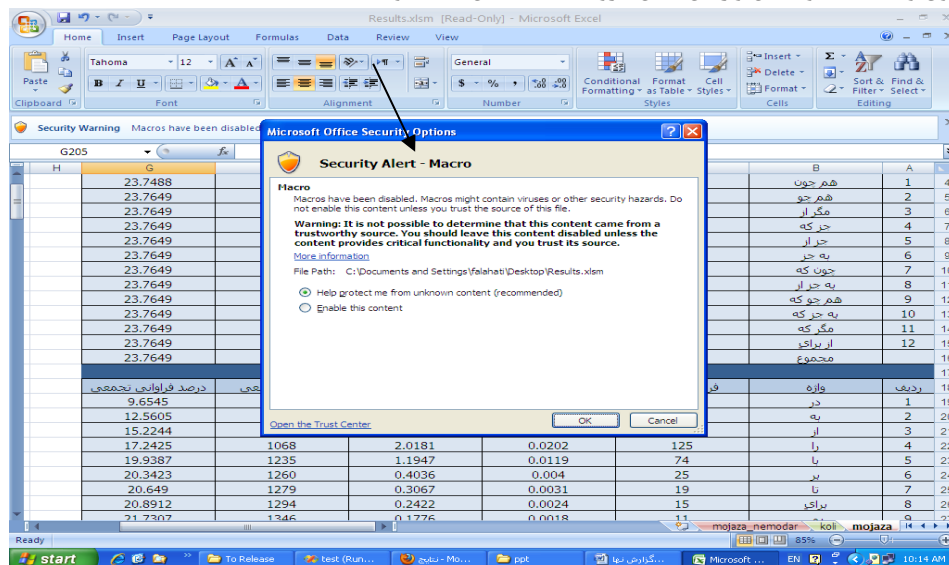
تصویر ۱۱: نمایش مرحله‌ی هشتم تحلیل متن در نرم‌افزار.

پس از کلید زدن بر روی دکمه‌ی «دریافت متن اکسل» صفحه‌ای مطابق تصویر ۱۱ ظاهر می‌شود. کارکرد این صفحه آن است که نتایج حاصله را در قالب یک فایل اکسل ذخیره می‌کند. پیش فرض اسم فایل "Results" است که می‌توان در تحلیل‌های مختلف و جهت تمایز نتایج مربوط به داده‌های متفاوت، به آن عددی مانند ۱، ۲ و ... را اضافه نمود. فایل حاصله را می‌توان با پسوند xism در پوشه مربوطه در رایانه‌ی شخصی ذخیره نمود.



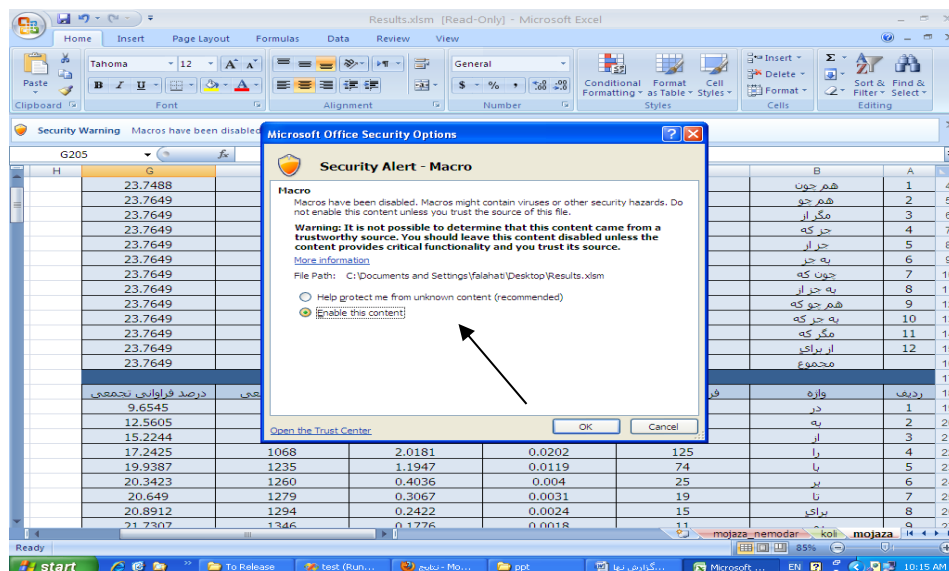
تصویر ۱۲: نمایش مرحله‌ی نهم تحلیل متن در نرم‌افزار.

پس از ذخیره‌ی نتایج در قالب فایل اکسل بر روی رایانه‌ی شخصی لازم است فایل مربوطه باز شود. فایل مربوطه اطلاعات آماری واژه‌های مربوط به هر یک از گروه‌های شش‌گانه را نمایش می‌دهد. همانگونه که پیشتر نیز ذکر شد، چنانچه اطلاعات تفکیکی مورد درخواست نبوده باشد و کاربر خواسته باشد درهم‌کرد اطلاعات مربوط به گروه‌های شش‌گانه را دریافت نماید، در تصویر ۱۲ به جای شش جدول، یک جدول واحد که حاوی کل واژگان شش گروه باشد، نمایش داده خواهد شد.



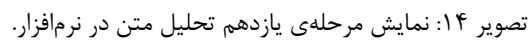
تصویر ۱۳-۱: نمایش مرحله‌ی دهم تحلیل متن در نرم‌افزار.

در بالای جدولی که در تصویر ۱۲ ارائه گردیده، گزینه‌ای تحت عنوان "Options" وجود دارد. برای اینکه نرم‌افزار بتواند کدهای مربوطه را به شکلی صحیح قرائت کند، لازم است بر روی این گزینه کلیک زده شود. با این کار، صفحه‌ی "Security Alert-Macro" باز می‌شود. پیش فرض در این گزینه، "Help Protect me from unknown content" می‌باشد. لازم است این پیش فرض به "Enable this content" تغییر یابد.

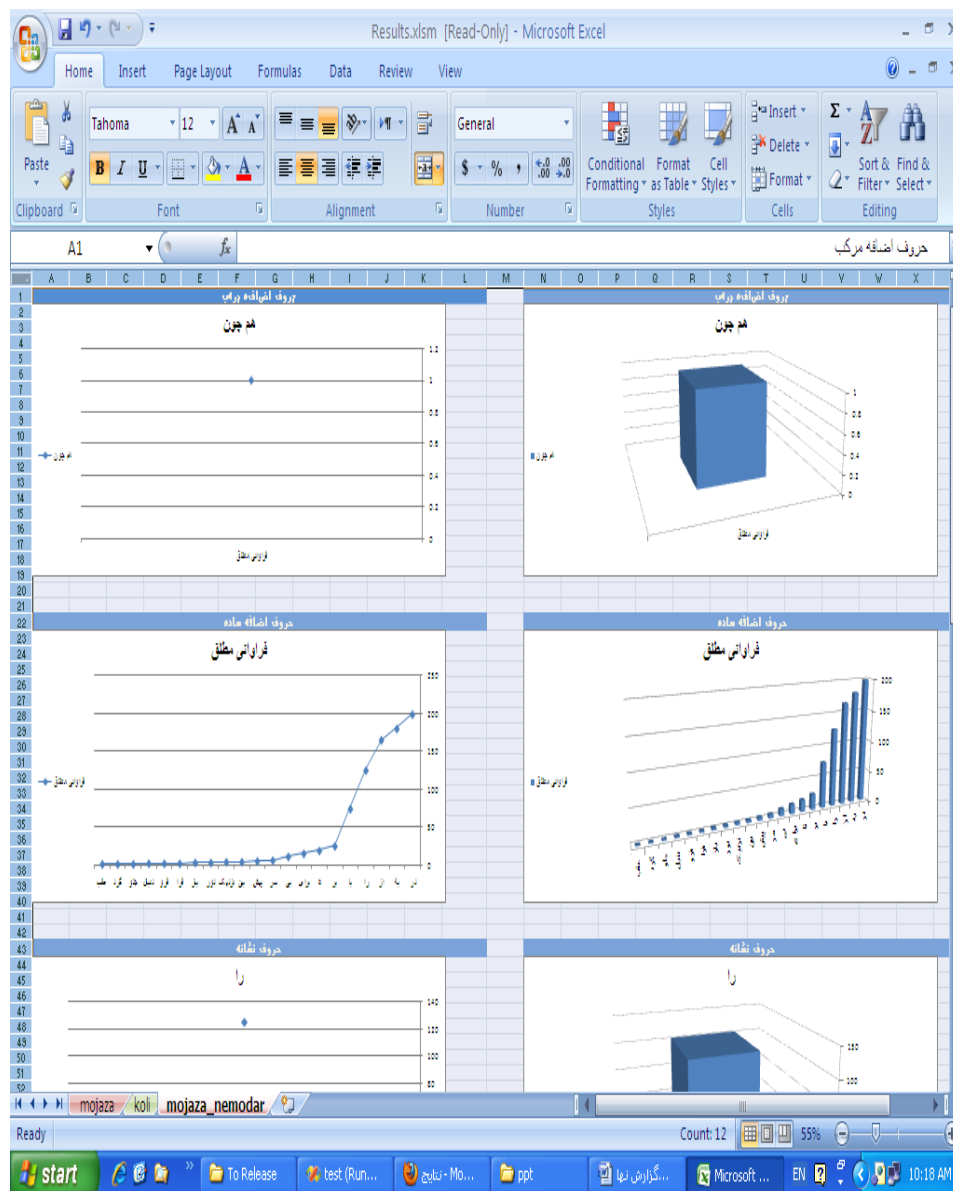


تصویر ۱۳-۲: نمایش مرحله‌ی دهم تحلیل متن در نرم‌افزار.

همانگونه که مشاهده می‌شود کاربر با تغییر پیش فرض، گزینه‌ی "Enable this content" را انتخاب کرده تا نرم‌افزار بتواند کلمات فارسی را به درستی شناسایی کرده، نمایش دهد.

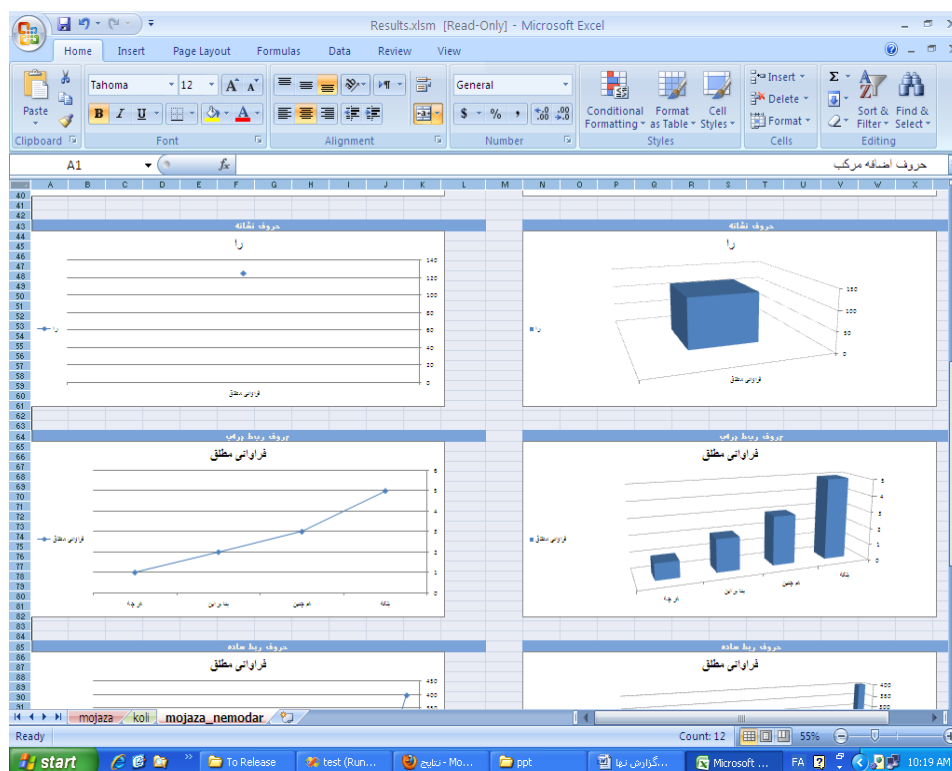


در این مرحله، چنانچه کاربر بخواهد از اطلاعات موجود در جداول استفاده نماید می‌تواند آنها را در فایل **Word** ذخیره نماید و در مقاله‌ی خود مورد استفاده قرار دهد. حال اگر مایل باشد از داده‌های خود نموداری نیز تهیه کند، می‌تواند بر روی گزینه‌ی «رسم نمودار» در تصویر ۱۴ کلیک بزند.

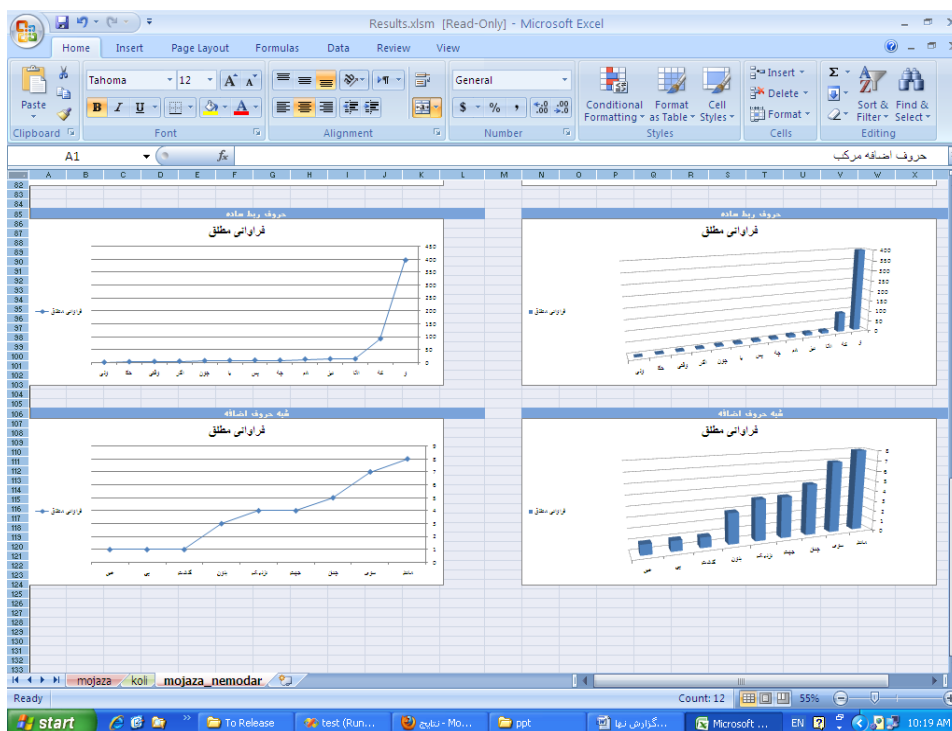


تصویر ۱۵-۱: نمایش مرحله‌ی دوازدهم تحلیل متن در نرم‌افزار (حروف اضافه مرکب و ساده).

پس از کلید زدن بر روی گزینه‌ی «رسم نمودار» در تصویر ۱۴، برای هر یک از گروه‌های شش‌گانه، دو نوع نمودار (یک نمودار ستونی و یک نمودار خطی) به شرح تصویرهای ۱۵-۱ الی ۱۵-۳ ترسیم می‌شود.



تصویر ۱۵-۲: نمایش مرحله‌ی دوازدهم تحلیل متن در نرم‌افزار (حروف نشانه و ربط مرکب).



تصویر ۱۵-۳: نمایش مرحله‌ی دوازدهم تحلیل متن در نرم‌افزار (حروف ربط ساده و شبه حروف اضافه).

۴- زمینه‌هایی برای مطالعات بیشتر

- در پژوهش حاضر صرفاً بر روی مقوله‌های بسته‌ای چون حروف (حروف اضافه، نشانه و ربط) کار شد. می‌توان نرم‌افزار مربوطه را در قالب پژوهشی دیگر به سایر حوزه‌ها و مقوله‌های دستوری گسترش داد.

- به دلیل پیچیدگی شناسایی، تنها حرف نشانه‌ی «را» از مجموعه‌ی حروف نشانه مورد پردازش قرار گرفت. تحلیل حروف نشانه‌ی دیگر نیز می‌تواند موضوع پژوهشی متمایز باشد. البته باید توجه داشت که هدف اولیه و مصوب پژوهش حاضر تحلیل حروف اضافه و ربط بسیط بوده که البته در جریان کار و به صورت داوطلبانه و با هدف ارتقاء کیفیت نرم‌افزار این دو حیطه به شش حیطه افزایش پیدا نمود.
- از طریق این نرم‌افزار و با تحلیل پیکره‌ی زبانی مناسب می‌توان حروف فعال و غیرفعال (مرده) را در زبان فارسی مشخص نمود. منظور از حروف فعال حروفی است که همچنان در فارسی امروز کاربرد دارد و مراد از حروف غیرفعال حروفی است که در فارسی امروز به کار برده نمی‌شوند.

۵- سخن پایانی

هدف از انجام پژوهش حاضر آن بود تا نرم‌افزاری تهیه شود تا بتوان به کمک آن، واژگان زبان فارسی متعلق به شش گروه حروف اضافه‌ی ساده، حروف اضافه مرکب، شبه حروف اضافه، حروف نشانه، حروف ربط و حروف ربط مرکب را شناسایی نمود و برای هر واژه اطلاعاتی آماری شامل فراوانی مطلق، فراوانی نسبی، درصد فراوانی نسبی، فراوانی تجمعی و درصد فراوانی تجمعی را در قالب جداول و نمودارهای خطی و ستونی ارائه کرد. برای انجام کار، فهرست واژگان مربوط به گروه‌های شش‌گانه فوق از کتاب‌های دستور زبان فارسی موجود در بازار استخراج شد. در مواردی که یک واژه در کتاب‌های مختلف به صورت‌های گوناگونی تحلیل و طبقه‌بندی شده بود، از نظر اکثریت استفاده شد و برای رفع مشکل تنوع املائی، برای هر واژه، تمامی صورت‌های املائی احتمالی، پیش‌بینی و فهرست شد. از معیار فاصله‌ی کامل قبل و بعد از واژه برای تفکیک واژه‌ها (به جز کسره‌ی اضافه) استفاده به عمل آمد و در شناسایی زنجیره‌ی حروف از قاعده‌ی بزرگترین طول استفاده شد. نرم‌افزار طراحی شده با درون‌داد متون نمونه مورد آزمایش قرار گرفت و خطاهای مشاهده شده، مرتفع گردید. این نرم‌افزار قادر است ضمن تحلیل اطلاعات، نتایج را در قالب فایل اکسل ارائه نماید. این کار، امکان استفاده از جداول و نمودارهای حاصله را به آسان‌ترین شکل در ساختار مقاله، کتاب و آثار دیگر فراهم می‌آورد و در تسریع نگارش مقالات و آثار علمی از سوی محققان حوزه‌ی زبان، ابزاری بسیار مناسب محسوب می‌شود. دستور نویسان و نحوین نیز می‌توانند از طریق تحلیل متون خود در نرم‌افزار مربوطه به کاربردی‌تر کردن دستورهای خود بیاورند. نتایج تحلیل متون در نرم‌افزار حاضر از سوی دو متخصص زبان مورد بررسی و بازبینی قرار گرفت و مطابقت نتایج با فهرست‌های واژگان طراحی شده، تأیید شد.

۶- تشکر و قدردانی

در اینجا لازم می‌داند از پروفسور تام کاب (Tom Cobb) که ایده‌ی اولیه‌ی طراحی نرم‌افزار حاضر را از تحقیقات ایشان در این زمینه در سایت لکس توتر گرفته ام تشکر نمایم. پیشتر نرم‌افزار ایشان را با استفاده از داده‌های انگلیسی ارزیابی کرده بودم و همان کار انگیزه‌ی اصلی شد برای طراحی نرم‌افزار حاضر در زبان فارسی.

فهرست منابع

- احمدی گیوی، ح. و انوری، ح. (۱۳۸۵). دستور زبان فارسی ۱ (ویرایش سوم). تهران: انتشارات فاطمی.
- شریعت، م. ج. (۱۳۸۴). دستور زبان فارسی (چاپ هشتم). تهران: انتشارات اساطیر.
- فرشید ورد، خ. (۱۳۸۲). دستور مفصل امروز (چاپ اول). تهران: انتشارات سخن.
- لازار، ژ. (۱۳۸۴). دستور زبان فارسی معاصر (ترجمه مهستی بحرینی). تهران: هرمس، مرکز بین المللی گفتگوی تمدن‌ها. (اثر اصلی در سال ۱۹۵۷ به چاپ رسید).
- مظفری، ز.، تاکتی، گ.، صباغ جعفری، م.، یوسفیان، پ. (۱۳۹۷). سامانه رفع ابهام معنایی از حروف اضافه در زبان فارسی با استفاده از قالب های معنایی. پژوهش های زبانی، ۹(۱)، ۹۹-۱۱۷.

- Alfawareh, H. M., & Jusoh, Sh. (2013). Resolving ambiguous preposition phrase for text mining applications. Retrieved January 21, 2020, from https://www.researchgate.net/publication/261384872_Resolving_ambiguous_preposition_phrase_for_text_mining_applications.
- Barbosa, J. J. G., Pazos, R. A., Gelbukh, A., Sidorov, G., Fraire-Huacuja, H. J., & Cruz, I. C. (2007). Prepositions and conjunctions in a natural language interfaces to databases. Retrieved February 21, 2020, from https://www.researchgate.net/publication/220945456_Prepositions_and_Conjunctions_in_a_Natural_Language_Interfaces_to_Databases.
- Cobb, T. (2019). Lextutor.ca.
- Yeh, A., & Vilain, M. B. (1998). Some properties of preposition and subordinate conjunction attachments. Retrieved on Jan. 12, 2020, from https://www.researchgate.net/publication/220483796_Some_Properties_of_Preposition_and_Subordinate_Conjunction_Attachments.