



مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری  
RICeST

## سامانه تحلیل حروف ربط و اضافه‌ی زبان فارسی

Persian Prepositions and Conjunctions Extractor and Analyzer

**دکتر محمدرضا فلاحتی قدیمی فومنی**

عضو هیأت علمی گروه پژوهشی زبانشناسی رایانه‌ای

تابستان ۱۳۹۲

سرشناسه	فلاحتی قدیمی فومنی، محمدرضا، ۱۳۴۵ -
عنوان و نام پدیدآور	سامانه‌ی تحلیل حروف ربط و اضافه‌ی زبان فارسی =
	/Persian prepositions and conjunctions extractor and...
	محمدرضا فلاحتی قدیمی فومنی؛ [به سفارش] مرکز منطقه‌ای اطلاع‌رسانی
	علوم و فناوری.
مشخصات نشر	شیراز: تخت جمشید، ۱۳۹۲.
مشخصات ظاهری	۶۴ ص.
شابک	978-600-283-069-2
وضعیت فهرست نویسی	فیپا
موضوع	فارسی -- حرف اضافه -- نرم‌افزار
موضوع	فارسی -- حرف ربط -- نرم‌افزار
شناسه افزوده	ایران. وزارت علوم، تحقیقات و فناوری. مرکز منطقه‌ای اطلاع‌رسانی علوم و
	فناوری
رده بندی کنگره	۱۳۹۲ س ۲ ب ۸ / ۲۷۹۱ PIR
رده بندی دیویی	۴۵۴
شماره کتابشناسی ملی	۳۲۸۱۳۷۰



## سامانه تحلیل حروف ربط و اضافه‌ی زبان فارسی

دکتر محمدرضا فلاحتی قدیمی فومنی

ناظر چاپ و آماده سازی کتاب: کریم فلاح □ صفحه‌آرایی و طرح جلد: نگار چراغی شیرازی  
لیتوگرافی و چاپ: پردیس و دنیا □ تیراژ: ۵۰۰ نسخه □ چاپ اول: تابستان ۱۳۹۲

مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری (RICEST)

روابط عمومی: ۰۷۱۱-۶۴۶۸۴۲۱ انتشارات: ۰۷۱۱-۶۴۶۸۴۵۲ نامبر: ۰۷۱۱-۶۴۶۸۳۵۲  
صفحه خانگی: [www.ricest.ac.ir](http://www.ricest.ac.ir) پست الکترونیک: [publication@ricest.ac.ir](mailto:publication@ricest.ac.ir)

انتشارات تخت جمشید

شیراز، خیابان پیروزی، تلفن: ۰۷۱۱-۲۲۴۵۴۰۱

## به نام خدا

نظر به اهمیت نقش اطلاع رسانی در زمینه‌های مختلف علوم و تکنولوژی، به منظور پاسخ‌گویی به نیازهای محققان و کارشناسان و ارائه تازه‌ترین اطلاعات علمی و فنی در کوتاه‌ترین زمان، مرکز منطقه‌ای اطلاع رسانی علوم و فناوری براساس تفاهم‌نامه منعقد شده میان وزارت علوم، تحقیقات و فناوری و فرهنگستان علوم جهان سوم در سال ۱۳۷۰ دایر و لایحه تاسیس آن در سال ۱۳۷۵ از تصویب مجلس شورای اسلامی و شورای نگهبان گذشت. این مرکز در چهارچوب ضوابط و مقررات آموزش عالی جمهوری اسلامی ایران فعالیت می‌نماید.

مرکز منطقه‌ای اطلاع رسانی علوم و فناوری از طریق توزیع آخرین اطلاعات علمی و فنی و نیز کمک به تامین منابع علمی مورد نیاز دانشگاهها، سازمانها و نهادهای تحقیقاتی و متخصصان داخلی و منطقه‌ای برای ارتقاء سطح علمی جمهوری اسلامی ایران و سایر کشورهای منطقه فعالیت می‌نماید.

کتاب «سامانه تحلیل حروف ربط و اضافه ی زبان فارسی» در راستای تحقق اهداف فوق تهیه شده و مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری افتخار دارد که شصت و سومین اثر خود را که حاصل تلاش فکری و طرح تحقیقاتی جناب آقای دکتر محمدرضا فلاحتی تقدیمی فومنی است به زیور طبع بیاراید و تقدیم جامعه علمی و فنی نماید.

امید است اقداماتی از این قبیل موجبات رضایت هرچه بیشتر جامعه علمی و فنی ایران را فراهم آورده و در راستای هدف ما که نشر و اشاعه اطلاعات علمی و فنی است، موثر واقع گردد.

**دکتر جعفر مهراذ**

**رئیس مرکز منطقه‌ای اطلاع رسانی علوم و فناوری**

**و سرپرست پایگاه استنادی علوم جهان اسلام (ISC)**

---

## اعضای شورای علمی

مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری (RICeST)

پایگاه استنادی علوم جهان اسلام (ISC)

---

دکتر جعفر مهران (رئیس شورا)

دکتر کرامت‌اله ایزدپناه - استاد دانشکده کشاورزی دانشگاه شیراز

دکتر عزت‌اله رئیسی استاد علوم زمین دانشکده علوم دانشگاه شیراز

دکتر علیرضا سپاسخواه - استاد دانشکده کشاورزی دانشگاه شیراز

دکتر هاجر ستوده - استادیار رشته علم اطلاعات و دانش‌شناسی دانشگاه شیراز

دکتر علی‌اکبر صفوی - استاد برق و الکترونیک دانشکده مهندسی دانشگاه شیراز

دکتر حمید علیزاده - استادیار مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری

دکتر محمد مهدی علیشاهی - استاد مکانیک دانشکده مهندسی دانشگاه شیراز

دکتر محمدرضا قانع - استادیار مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری

دکتر علی گزنی - استادیار علم اطلاعات و دانش‌شناسی پایگاه استنادی علوم جهان

اسلام

دکتر جلیل مقدسی - استاد شیمی دانشکده علوم دانشگاه شیراز

دکتر محمد نمازی - استاد حسابداری دانشکده علوم اجتماعی دانشگاه شیراز

## تقدیر و قدردانی:

در اینجا لازم می‌دانم از زحمات تمامی کسانی که به نحوی به اینجانب در انجام پژوهش حاضر یاری رساندند تشکر نمایم. در آغاز از ریاست محترم مرکز منطقه‌ای اطلاع رسانی علوم و فناوری جناب آقای پروفیسور دکتر جعفر مهرداد و هیات رئیسه محترم که با تصویب پژوهش حاضر زمینه انجام آن را فراهم نمودند سپاس‌گزاری می‌نمایم. دوست و همکار بسیار بزرگواریم آقای شاپوررضا برنجیان نیز در جای خود شایسته قدردانی است. مطمئناً بدون حمایت و نظرات سازنده ایشان پژوهش حاضر بدین سان ارایه نمی‌گردید. از سرکار خانم مهندس حسینیا نیز که کار برنامه‌نویسی مربوط به طرح را با نظم و دقتی خاص به انجام رساندند تشکر می‌نمایم. از سرکار خانم روزی‌طلب نیز که زحمت تایپ گزارش نهایی طرح را تقبل فرمودند سپاس گزارم. در پایان، لازم می‌دانم از تمامی همکاران و دوستانی که به دلیل محدودیت فضا نامشان در این مجموعه آورده نشده است به دلیل عدم ذکر نامشان عذرخواهی و به دلیل کمک فکری شان تقدیر و قدردانی نمایم.

دکتر محمد رضا فلاحتی قدیمی فومنی

مرداد ۱۳۹۲



## چکیده:

امروزه انجام پردازش‌ها و تحلیل‌های زبانی با استفاده از نرم افزار 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 شش گانه فوق از کتاب‌های دستور زبان فارسی موجود در بازار استخراج گردید. در مواردی که یک واژه در کتاب‌های مختلف به صورت‌های گوناگونی تحلیل و طبقه بندی شده بود، از نظر اکثریت استفاده شد و برای رفع مشکل تنوع املائی، برای هر واژه، تمامی صورت‌های املائی احتمالی، پیش بینی 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 نرم افزار 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 نحویان 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، نرم‌افزار تحلیل زبانی، دستورنویسی خودکار،

حروف اضافه، حروف ربط، دستور زبان فارسی، تحلیل بسامدی.

## فهرست مندرجات

---

صفحه	عنوان
أ	تقدير و قدردانی
ب	چکیده
ج	فهرست مندرجات
ه	فهرست تصاویر
ا	پیشگفتار
٤	مقدمه
٩	الگوریتم برنامه
٩	دریافت متن
٩	تبدیل فایل به فرمت .txt
١٠	سیاهه کلمات
١٠	تحلیل متن
١٢	تعیین نوع مرتب سازی خروجی
١٢	اجزای خروجی و نحوه نمایش
١٢	تولید جداول و نمودار در قالب فایل اکسل
١٢	چند نکته دیگر
١٥	شناسایی و رفع خطاهای برنامه



## فهرست مندرجات

---

صفحه	عنوان
۱۶.....	ارزیابی برنامه
۱۷.....	توصیف گام به گام نرم افزار
۴۱.....	زمینه‌هایی برای مطالعات بیشتر
۴۲.....	خلاصه
۴۳.....	فهرست منابع
	پیوست ۱: زنجیره املائی حروف اضافه ساده و مرکب
۴۴.....	جهت رفع مشکل تنوع املائی
۴۵.....	پیوست ۲: زنجیره املائی شبه حروف اضافه جهت رفع مشکل تنوع املائی
	پیوست ۳: زنجیره املائی حرف نشانه و حروف ربط ساده
۴۶.....	جهت رفع مشکل تنوع املائی
	پیوست ۴: زنجیره املائی حروف ربط مرکب
۴۷.....	جهت رفع مشکل تنوع املائی

## فهرست تصاویر

---

عنوان	صفحه
تصویر ۱: نمایش صفحه ی اصلی نرم افزار .....	۱۷
تصویر ۲: نمایش جعبه ی ارسال پیشنهاد برای سامانه .....	۱۸
تصویر ۳: نمایش مرحله ی اول تحلیل متن در نرم افزار .....	۱۹
تصویر ۴: نمایش مرحله ی دوم تحلیل متن در نرم افزار .....	۲۰
تصویر ۵: نمایش مرحله ی دوم تحلیل متن در نرم افزار (ادامه تصویر ۴) .....	۲۱
تصویر ۶: نمایش مرحله ی سوم تحلیل متن در نرم افزار .....	۲۲
تصویر ۷-۱: نمایش مرحله ی چهارم تحلیل متن در نرم افزار (جستجوی فایل) .....	۲۳
تصویر ۷-۲: نمایش مرحله ی چهارم تحلیل متن در نرم افزار (انتخاب فایل) .....	۲۳
تصویر ۸: نمایش مرحله ی پنجم تحلیل متن در نرم افزار .....	۲۴
تصویر ۹: نمایش مرحله ی ششم تحلیل متن در نرم افزار .....	۲۴
تصویر ۱۰-۱: نمایش مرحله ی هفتم تحلیل متن در	
نرم افزار (نمایش نتایج به صورت کلی) .....	۲۵
تصویر ۱۰-۲: نمایش مرحله ی هفتم تحلیل متن در	
نرم افزار (نمایش نتایج به صورت مجزا) .....	۲۶
تصویر ۱۱: نمایش مرحله ی هشتم تحلیل متن در نرم افزار .....	۲۷
تصویر ۱۲: نمایش مرحله ی نهم تحلیل متن در نرم افزار .....	۲۸

## فهرست تصاویر

---

عنوان	صفحه
تصویر ۱۳-۱: نمایش مرحله‌ی دهم تحلیل متن در نرم افزار	۲۹
تصویر ۱۳-۲: نمایش مرحله‌ی دهم تحلیل متن در نرم افزار (ادامه تصویر ۱۳-۱).....	۳۰
تصویر ۱۴: نمایش مرحله‌ی یازدهم تحلیل متن در نرم افزار	۳۱
تصویر ۱۵-۱: نمایش مرحله‌ی دوازدهم تحلیل متن در نرم افزار (حروف اضافه مرکب و ساده)	۳۲
تصویر ۱۵-۲: نمایش مرحله‌ی دوازدهم تحلیل متن در نرم افزار (حروف نشانه و ربط مرکب)	۳۳
تصویر ۱۵-۳: نمایش مرحله‌ی دوازدهم تحلیل متن در نرم افزار (حروف ربط ساده و شبه حروف اضافه)	۳۴
تصویر ۱۶-۱: نمایش اطلاعات مربوط به حروف اضافه‌ی مرکب (نمودار خطی)	۳۵
تصویر ۱۶-۲: نمایش اطلاعات مربوط به حروف اضافه‌ی مرکب (نمودار ستونی)	۳۵
تصویر ۱۶-۳: نمایش اطلاعات مربوط به حروف اضافه‌ی ساده (نمودار خطی)	۳۶

## فهرست تصاویر

---

عنوان	صفحه
تصویر ۱۶-۴: نمایش اطلاعات مربوط به حروف	
اضافه ساده (نمودار ستونی) .....	۳۶
تصویر ۱۶-۵: نمایش اطلاعات مربوط به حرف	
نشانه "را" (نمودار خطی) .....	۳۷
تصویر ۱۶-۶: نمایش اطلاعات مربوط به حرف	
نشانه "را" (نمودار ستونی) .....	۳۷
تصویر ۱۶-۷: نمایش اطلاعات مربوط به حروف	
ربط مرکب (نمودار خطی) .....	۳۸
تصویر ۱۶-۸: نمایش اطلاعات مربوط به حروف	
ربط مرکب (نمودار ستونی) .....	۳۸
تصویر ۱۶-۹: نمایش اطلاعات مربوط به حروف	
ربط ساده (نمودار خطی) .....	۳۹
تصویر ۱۶-۱۰: نمایش اطلاعات مربوط به حروف	
ربط ساده (نمودار ستونی) .....	۳۹
تصویر ۱۶-۱۱: نمایش اطلاعات مربوط به شبه حروف اضافه (نمودار خطی).....	۴۰

## فهرست تصاویر

---

صفحه

عنوان

تصویر ۱۶-۱۲: نمایش اطلاعات مربوط به شبه حروف

۴۰ ..... اضافه (نمودار ستونی)



## پیشگفتار:

امروزه بسیاری از امور پژوهشی زمان بر - که زمانی صرفاً توسط انسان انجام می‌شد - به ماشین سپرده شده که این امر خود به روند انجام پژوهش و امور در جوامع انسانی شتاب بیشتری داده است. هدف از پژوهش حاضر آن بوده تا جهت استخراج و تحلیل خودکار حروف اضافه، نشانه و ربط فارسی نرم افزاری تهیه شود. نرم افزاری که بتواند به محققان حوزه‌ی زبان در نگارش مقالات علمی و تحلیل آماری مقوله های دستوری یاری برساند.

ایده‌ی اولیه‌ی طراحی چنین نرم افزاری به سال ۱۳۸۶ باز می‌گردد. در آن سال با مجموعه‌ای از نرم افزارها که توسط پروفسور تام کاب (Tom Cobb) طراحی و ارائه گردیده بود آشنا شدم. البته، آن نرم افزار صرفاً مقوله‌هایی را برای تمام واژگان محاسبه می‌نمود، حال آنکه در نرم افزاری که در تحقیق حاضر ارائه می‌گردد بر روی مقوله‌های بسته‌ی دستوری تمرکز گردیده (که این امر خود بر دقت نرم افزار می‌افزاید) ضمن آنکه قابلیت تولید نمودار و جداول در قالب فایل اکسل را نیز دارد که در نرم افزار ارائه شده توسط پروفسور تام کاب این قابلیت وجود نداشت. آن سال، سال آغازین دوره‌ی دکتری اینجانب در حوزه‌ی زبان شناسی رایانه‌ای نیز بود. یکی از نرم افزارهایی که توجه‌ام را بسیار به خود جلب کرد نرم افزاری بود که واژگان موجود در متون انگلیسی را از نظر فراوانی رخداد مورد بررسی قرار می‌داد. آن نرم‌افزار را بر روی فارسی امتحان کردم که حاصل آن مقاله‌ای زبانشناختی بود که در سال ۲۰۰۸ با عنوان:

*Can an English word frequency counter meet the requirements of Persian automatic indexing?*

در شماره ۱ از جلد ۲ مجله *International Journal of Theoretical Linguistics*

به چاپ رسید.

در آن پژوهش دریافتیم که نرم افزارهای تحلیل آماری به خاطر ترکیبات خاص واژگانی و قواعد واژه سازی موجود در فارسی در تحلیل بسامدی واژگان فارسی با مشکلات بسیاری مواجه‌اند. از اینرو درصدد برآمدیم در مرحله‌ی اول مقوله‌هایی را موضوع طرح قرار دهیم که استخراج آن‌ها از پیچیدگی بسیار بالایی برخوردار نباشد. به همین خاطر، مقوله‌های بسته‌ای چون حروف اضافه، نشانه و ربط برگزیده شد. هدف اولیه، پرداختن به حروف اضافه و ربط ساده بود که در جریان کار این دو مقوله به شش مقوله گسترش یافت که عبارت بودند از: (۱) حروف اضافه ساده، (۲) حروف اضافه مرکب، (۳) شبه حروف اضافه، (۴) حرف نشانه "را"، (۵) حروف ربط ساده و (۶) حروف ربط مرکب. درست به همین خاطر عنوان پژوهش از "استخراج و تحلیل خودکار حروف اضافه و ربط بسیط در متون فارسی" به "سامانه تحلیل حروف ربط و اضافه‌ی زبان فارسی" تغییر یافت (البته، با توجه به این که تنها حرف نشانه "را" در این پژوهش مورد تحلیل قرار گرفت، از به کارگیری لفظ "حرف نشانه" در عنوان کتاب حاضر و نیز در نرم افزار تهیه شده خودداری گردید، هرچند در عمل نرم افزار مربوطه این حرف نشانه را به درستی تحلیل می‌کند).

البته، در استخراج اعضای هر مقوله مشکلات بسیاری نیز وجود داشت. بدین ترتیب که در کتاب‌های مختلف دستور فارسی نحوه‌ی تقسیم‌بندی این گروه‌ها گاه مبهم است و بین گروه‌ها همپوشانی نیز وجود دارد. نویسندگان مختلف گاه یک واژه‌ی خاص را در مقوله‌های مختلف قرار می‌دهند. این گونه مسایل سبب گردید تا تهیه فهرست واژگان چندان هم که تصور می‌شد ساده نباشد. سعی شد با بررسی کتب مختلف در مواردی که برای ضبط یک واژه در یک گروه مشکل وجود داشت، از نظر اکثریت استفاده شود و اگر در بیشترین کتب یک واژه مثلاً حرف اضافه‌ی ساده قلمداد گردیده بود، آن واژه در گروه حروف اضافه‌ی ساده قرار گیرد. در مواردی که همچنان با مشکل مواجه بودیم (البته این موارد انگشت شمار بود) از نظر متخصصین زبان و دستور فارسی جهت ثبت واژه در یک گروه خاص استفاده شد.

در تهیه این نرم افزار سعی گردید تا مشکلاتی چون نوع فاصله (کامل، نیم، بی‌فاصلگی) مسئله ساز نشود. به بیان دیگر، نرم افزار صورت‌های زبان را با هر نوع



فاصله‌ای که نوشته شده باشند شناسایی می‌کند، هر چند برای نمایش نتیجه از یک صورت ظاهری صرف استفاده شده است.

ضمناً، از آنجا که هدف نهایی نرم افزار مربوطه کمک به تولید و افزایش سرعت و نگارش مقالات و آثار علمی در حوزه‌ی دستور زبان فارسی بوده است، نرم افزار به گونه‌ای طراحی شده که می‌تواند در قالب فایل اکسل جداول و نمودارهایی را نیز ارائه نماید که هر یک نحوه‌ی ظهور حروف اضافه، نشانه و ربط را در متن مربوطه نشان می‌دهند.

این نرم افزار قادر است حجم بزرگی از فایل را با پسوند doc یا docx و یا txt دریافت دارد و آن را مورد تحلیل قرار دهد. می‌توان اطلاعات آماری را برای هر یک از گروه‌های شش‌گانه به صورت مجزا یا در هم کرد دریافت نمود.

دو نوع نمودار خطی و ستونی نیز موجود است که کاربر براساس نیاز خاص خود می‌تواند یکی را برگزیند.

این پروژه درگام‌های بعدی به دیگر مقوله‌های زبان فارسی نیز خواهد پرداخت و امید می‌رود بتواند به عنوان یک ابزار جانبی به محققان حوزه‌ی زبان در تولید انتشارات علمی یاری برساند.

در کار انجام شده ممکن است نواقصی نیز وجود داشته باشد که امیدواریم با بررسی و ارزیابی عملی، کاربران محترم ما را از نظرات سازنده‌ی خود جهت بهبود برنامه مطلع فرمایند.

## مقدمه:

در زبان 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 زاید 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. البته، این دسته از کلمات در آغاز معنی خاص و مستقلی داشته‌اند و بر اثر استعمال در مورد ثانوی بوده است که به تدریج معنی اصلی خود را از دست داده‌اند (همان، ص. ۳۱۱). 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 نشان می‌دهد که دست‌نویسان فارسی به حروف نگاهی کاملاً یکدست نداشته‌اند. در حقیقت نمی‌توان یک تقسیم بندی واحد را در تمامی آثار مشاهده نمود، ضمن آنکه در خصوص عضویت کلمات در یک گروه خاص نیز اختلاف نظر بسیار وجود دارد. 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 گردیده است.

هدف اصلی و اولیه از پژوهش حاضر پرداختن به دو مقوله‌ی "حروف اضافه‌ی ساده" و "حروف ربط ساده" بوده است. البته، در ضمن کار و پیرو بررسی کتاب‌های مختلف دستوری مانند دستور زبان فارسی (شریعت، ۱۳۸۴)، دستور زبان فارسی گیوی و انوری (۱۳۸۸)، دستور مفصل امروز (فرشید ورد، ۱۳۸۲) و دستور زبان فارسی معاصر (لازار، ۱۳۱۷) که این اثر در ۱۳۸۴ توسط خانم بحرینی به فارسی ترجمه گردید، تقسیم‌بندی‌های دیگری نیز مشاهده شد و دو مقوله‌ی اولیه‌ی حروف اضافه‌ی

ساده و حروف ربط ساده به مقوله‌های شش‌گانه‌ی حروف اضافه‌ی ساده، حروف اضافه‌ی مرکب، شبه حروف اضافه‌ها، حروف ربط ساده، حروف ربط مرکب و حرف نشانه "را" گسترش یافت. در ادامه، هر یک از مقوله‌های شش‌گانه فوق به اختصار معرفی می‌گردد:

#### ۱- حروف اضافه

حروف اضافه کلماتی هستند که قبل از اسم یا جانشین آن می‌آیند و اسم 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 نکات به الا، ایدون، با، باز، تا، جز، چون، که، مگر و درنهایت و اشاره می‌کند. (ص. ۳۱۲)

#### ۲- حروف اضافه‌ی مرکب

حرف اضافه‌ی مرکب آن است که بیش از یک حرف باشد. حرف‌های وابستگی مرکب بیشتر از بهم پیوستگی 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 اضافه‌ی مرکب است، حال آنکه برخی دیگر از دستورنویسان آن را حرف اضافه‌ی ساده محسوب می‌دارند و کسره‌ی اضافه را واحدی مستقل فرض نمی‌کنند.

### ۳- شبه حروف اضافه

شبه حروف اضافه از به هم پیوستن یک یا دو حرف اضافه‌ی ساده با یک اسم و گاهی با یک صفت یا یک ضمیر اشاره ساخته می‌شوند. شبه حرف اضافه، عمل یک حرف اضافه‌ی ساده را با صراحت و دقت بیشتر انجام می‌دهد. شریعت (۱۳۸۴) در تعیین ملاکی برای شناختن شبه حرف اضافه بیان می‌دارد که شبه حرف اضافه آن است که بتوان آن را حذف کرد و به جای آن یک حرف اضافه‌ی 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، بیاب، به جای، به جهت، بدون، برای، بسان، به سر، به سوی، بغیر، به کردار، به نزدیک، با سر، برجای، برسان، بر سر، بر کردار، درباب، در جنب، در حق (هه به کسر آخر) چنانچون (ص ۳۱۵).

البته، همان گونه که پیشتر نیز ذکر شد دستوریان مختلف حروف و به طور اخص حروف اضافه را به اشکال مختلف تقسیم می‌کنند. به عنوان نمونه، باطنی (۱۳۷۲) حروف اضافه‌ی فارسی را به دو دسته تقسیم می‌کند: دسته‌ای که اضافه می‌پذیرند و دسته‌ای که اضافه نمی‌پذیرند.

#### ۴- 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، هرکجا، چندان که، هرچند 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 "دیدی" 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- حروف ربط مرکب

حرف ربط را از نظر ساختمان به دسته‌های مختلفی تقسیم می‌کنند که عبارتند از:

الف- حروف ربط ساده که بیشتر 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: همیکنه، 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 (همان، ص. ۳۲۳).

#### ۶- حروف نشانه

نشانه‌ها در فارسی متعددند اما برای پرهیز از ابهام و امکان تحلیل آماری بدون تحلیل‌های معناساختی پیچیده صرفاً حرف نشانه‌ی "را" در پژوهش حاضر مورد بررسی قرار گرفت.

## الگوریتم برنامه

در این بخش، زیر ساخت نرم افزار به ترتیب مراحل اجرایی تشریح خواهد شد.

### ۱- دریافت متن.

نرم افزار حاضر به گونه‌ای طراحی گردیده است که می‌تواند فایل‌های دارای پسوند `.doc`، `.docx` و `.txt` را دریافت نماید. البته، در نرم افزار جعبه‌ای طراحی شده که می‌توان بخشی از فایل متنی را از روی رایانه‌ی شخصی کپی و در نرم افزار وارد نمود.

### 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 می‌گردد. بدین ترتیب، تمامی نمودارها و اطلاعات فرامتنی از متن کنار گذاشته می‌شود و تنها متن مربوطه باقی می‌ماند. اطلاعات باقی مانده، با پسوند `.txt` ذخیره می‌گردد.

### ۳- سیاهه کلمات

با توجه به اینکه در نرم افزار حاضر شش گروه کلمات شامل حروف اضافه‌ی ساده، حروف اضافه‌ی مرکب، شبه حروف اضافه، حروف ربط ساده، حروف ربط مرکب و حرف نشانه "را" مورد پردازش قرار می‌گیرند، فهرستی از اعضای هر گروه توسط محقق و با استفاده از کتاب‌های دستور زبان فارسی تهیه گردید. این اطلاعات به عنوان "سیاهه کلمات" در قالب شش فهرست جدا و یک فهرست درهم کرد تهیه گردید و در ساختار نرم افزار قرار گرفت. با توجه به وجود انواع مختلف فاصله‌ها و شیوه‌های مختلف پیوند واژه در فارسی، تمامی صورت‌های احتمالی هر واژه توسط محقق پیش بینی و فهرست گردید. مزیت این کار آن است که می‌توان تمامی رخدادهای یک واژه را (با هر صورت ظاهری که نوشته شده باشد) شناسایی، فهرست و درهم کرد نمود. البته، اگرچه صورت‌های مختلف شناسایی می‌گردد، تنها از یک صورت ظاهری به عنوان نشان دهنده‌ی مدخل که حاوی درهم کرد اطلاعات آماری اشکال مختلف آن واژه است، در بخش تحلیل اطلاعات استفاده می‌شود. بدین ترتیب، "از این رو" و "از اینرو" هر دو مورد شناسایی قرار می‌گیرند. منتهی برای نمایش نتیجه‌ی نهایی تنها از یکی از این دو صورت استفاده می‌شود و آماری که برای این صورت منتخب ارایه می‌گردد در بردارنده‌ی آماری است که برای تمامی صورت‌های ظاهری مختلف موجود در متن بدست آمده است.

### ۴- تحلیل متن

تمامی واژه‌های موجود در سیاهه کلمات در فایل متنی جستجو می‌شوند و برای هر واژه پنج پارامتر محاسبه می‌گردد که عبارتند از: (الف) فراوانی مطلق، (ب) فراوانی نسبی، (ج) درصد فراوانی نسبی، (د) فراوانی تجمعی و (ه) درصد فراوانی تجمعی. شیوه‌ی محاسبه‌ی متغیرهای فوق به شرح زیر است:

الف- فراوانی مطلق: این مختصه تعداد دفعاتی را نشان می‌دهد 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 می‌شود. به عنوان مثال، اگر واژه‌ی "به" ۱۰ بار در متنی که حاوی ۱۰۰ واژه بوده ظاهر شود، فراوانی نسبی این واژه در متن برابر با ۱/ خواهد بود.

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

د- *فراوانی تجمعی*: عدد این مختصه در پله‌ی اول همان عدد فراوانی مطلق است. برای محاسبه‌ی فراوانی تجمعی در پله‌ی دوم، فراوانی مطلق در پله‌ی اول و دوم با هم جمع می‌شود. در پله‌ی سوم، فراوانی تجمعی از طریق افزودن فراوانی مطلق در پله‌های یک، دو و سه بدست می‌آید و به همین شکل تا آخر. فراوانی تجمعی میزان حجمی را که واژگانی خاص در یک متن به خود اختصاص می‌دهند نشان می‌دهد.

ه- *درصد فراوانی تجمعی*: این مختصه در پله‌ی اول از طریق ضرب فراوانی نسبی پله‌ی مربوطه در ۱۰۰ بدست می‌آید. مثلاً اگر حرف "و" فراوان ترین واژه و دارای فراوانی نسبی 0.0453 بوده باشد، درصد فراوانی تجمعی در همین پله معادل 4.503 خواهد بود. حال اگر واژه‌ی ردیف دو "به" و فراوانی نسبی آن 0.0237 باشد درصد فراوانی تجمعی برای این دو پله از طریق ضرب فراوانی نسبی در هر پله در ۱۰۰ و سپس جمع نتایج محاسبه می‌شود.

$$\text{پله ۱: "و" فراوانی نسبی } 0.0453 \text{ درصد فراوانی تجمعی } 4.53$$

$$\text{پله ۲: "به" فراوانی نسبی } 0.0237 \text{ درصد فراوانی تجمعی } 2.37 + 4.53$$

$$\text{پله ۳: "در" فراوانی نسبی } 0.0237 \text{ درصد فراوانی تجمعی } 2.37 + 2.37 + 4.53$$

### ۵- تعیین نوع مرتب سازی خروجی

پیش از تحلیل اطلاعات، کاربر می‌تواند نوع مرتب سازی خروجی را تعیین نماید. در نرم افزار حاضر، سه نوع مرتب سازی وجود دارد که کاربر براساس نیاز خاص خود می‌تواند یکی را برگزیند. این سه نوع مرتب سازی عبارتند از: (الف) براساس حروف الفبا، (ب) براساس فراوانی مطلق به صورت صعودی و (ج) براساس فراوانی مطلق بصورت نزولی.

### ۶- اجزای خروجی و نحوه‌ی 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. نحوه‌ی نمایش خروجی نیز به دو شیوه می‌باشد. در شیوه‌ی اول، تمامی اطلاعات مربوط به گروه‌های شش گانه (حروف اضافه‌ی ساده، حروف اضافه‌ی مرکب، شبه حروف اضافه، حرف نشانه "را"، حروف ربط ساده و حروف ربط مرکب) به صورت درهم کرد و در قالب جدولی واحد ارائه می‌شود. در شیوه‌ی دوم، اطلاعات مربوط به هر یک از گروه‌های شش گانه در قالب جدولی مجزا ارائه می‌گردد.

### ۷- تولید جداول و نمودار در قالب فایل اکسل

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

### ۸- چند نکته‌ی دیگر

الف- به عنوان بخشی از عملیات پیش پردازش، 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 استفاده از علائم نگارشی و

نشانه‌گذاری نیز از هم تفکیک می‌شوند، در مرحله‌ی پیش پردازش، علائم نشانه‌گذاری زیر با فاصله‌ی کامل جایگزین شد:

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

ب- همان گونه 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 پیش از 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، نرم افزار توانست واژه‌ی "برای" را ترکیبی از "برای + کسره‌ی اضافه" شناسایی نماید.

ج- شناسایی و کنترل صورت‌های مختلف نگارشی. 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 "بجای"؛ "هم چنین" در مقابل "همچنین")، نرم افزار به گونه‌ای طراحی گردید که بتواند این گونه‌های صوری مرتبط را در قالب مدخلی واحد تحلیل و گزارش نماید. تنوع نگارش ناشی از اتصال و انفصال و نیز بکارگیری انواع مختلف فاصله و یا نشانه‌های زبر زنجیری از سوی محقق برای هر واژه فهرست گردید و در واقع برای هر واژه تمامی حالت‌های مختلف احتمالی که واژه می‌توانست با آن در متن ظاهر شود، مشخص شد. در گام بعدی، نرم افزار پس از شناسایی هر صورت و ارزیابی اطلاعات آماری آن و پیش از ارزیابی 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،

نرم افزار یکی از این دو صورت را به عنوان واژه‌ی مدخل برمی‌گزیند و عدد ۹ را مقابل آن یادداشت می‌کند.

د- با توجه به اینکه در گروه‌های شش گانه مورد بررسی، حروف بسیط و مرکب وجود داشت، لازم بود درخصوص 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" در نظر گرفته شد. به بیان ساده‌تر، در تحلیل عبارت "از برای" نرم افزار ابتدا "از" را تشخیص می‌دهد. بلافاصله "برای" نیز که پیش و پس از آن فاصله‌ی کامل آمده، شناسایی می‌گردد. با استفاده از قاعده‌ی تعریف شده، نرم افزار به جای اینکه "از" و "برای" را دو مدخل فرض کند، به دلیل اینکه واژه‌ای 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 واژه‌ها همواره رویت بزرگترین طول عبارت خواهد بود.

ه- سرعت پردازش اطلاعات در نرم افزار تا حد زیادی به RAM و قدرت رایانه‌ی شخصی بستگی دارد. در یک رایانه‌ی معمولی با RAM معادل 2GB این نرم افزار می‌تواند تمامی 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- شناسایی و رفع خطاهای برنامه

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

الف- 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.

ب- از آنجا که واژگان گروه‌های شش گانه 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 شده بود، با مشورت یک متخصص زبان و جستجوی نظر غالب در کتاب‌های دستور زبان فارسی و 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.

ج- بخشی تحت عنوان "ارسال پیشنهاد" به نرم‌افزار اضافه گردید.

د- 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- ارزیابی برنامه**

دو متخصص زبان با مدرک کارشناسی ارشد زبان و ادبیات فارسی از طریق درونداد متونی کوتاه به برنامه نتایج را با لیست‌های شش‌گانه‌ی تولید شده مطابقت دادند و صحت عملکرد نرم‌افزار را از طریق مقایسه‌ی تحلیل به صورت دستی با تحلیل خودکار تایید نمودند.

## توصیف گام به گام نرم افزار



امروزه بسیاری از امور پژوهشی زمان بر - که زمانی صرفاً توسط انسان انجام می شد- به ماشین سپرده شده که این امر خود به روند انجام پژوهش و امور در جوامع انسانی شتاب بیشتری داده است. هدف از پژوهش حاضر آن بوده تا جهت استخراج و تحلیل خودکار حروف اضافه و ربط فارسی نرم افزاری تهیه شود. نرم افزاری که بتواند به محققان حوزه ی زبان در نگارش مقالات علمی و تحلیل آماری مقوله های دستوری باری برساند.

ایده ی اولیه ی طراحی چنین نرم افزاری به سال 1386 باز می گردد. در آن سال با مجموعه ای از نرم افزارها که توسط پروفسور تام کاب طراحی و ارائه گردیده بود آشنا شدم. این سال، سال آغازین دوره ی دکتری اینجانب در حوزه ی زبان شناسی رایانه ای نیز بود. یکی از نرم افزارهایی که توجه ام را بسیار به خود جلب کرد نرم افزاری بود که واژگان موجود در متون انگلیسی را از نظر فراوانی رخداد مورد بررسی قرار می داد. آن نرم افزار را برای فارسی امتحان کردم که حاصل آن مقاله ای بود که در سال 2008 با عنوان:

Can an English word frequency counter meet the requirements of Persian automatic indexing?  
International Journal of Theoretical Linguistics, Vol.2, No.1, PP.18-29 به چاپ رسید.

در آن پژوهش دریافتم که نرم افزارهای تحلیل آماری به خاطر ترکیبات خاص واژگانی و قواعد واژه سازی موجود در فارسی در تحلیل بسامدی واژگان فارسی با مشکلات بسیاری مواجه اند. از این رو درصدم برآمدم در مرحله ی اول مقوله هایی را موضوع طرح قرار دهم که استخراج آنها از پیچیدگی بسیار بالایی برخوردار نباشد. از این رو مقوله های بسته ای چون حروف اضافه و حروف ربط انتخاب شد. هدف اولیه برداختن به حروف اضافه و ربط ساده بود که در جریان کار این دو مقوله به شش مقوله گسترش یافت که عبارت بودند از: (1) حروف اضافه ساده، (2) حروف اضافه مرکب، (3) شبه حروف اضافه، (4) حروف نشانه، (5) حروف ربط ساده و (6) حروف ربط مرکب. درست به همین خاطر عنوان نرم افزار از "استخراج و تحلیل خودکار حروف اضافه و ربط بسیط در متون فارسی" به "سامانه تحلیل حروف ربط، نشانه

### تصویر ۱: نمایش صفحه ی اصلی نرم افزار

تصویر ۱ نمایش کلی از صفحه اصلی سامانه را نشان می دهد. این صفحه حاوی سه بخش عمده است که عبارتند از: ۱- صفحه اصلی، ۲- تحلیل متن و ۳- ارسال

پیشنهاد. 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 پیشنهاد نرم افزار را نشان می‌دهد. کاربران پس از درونداد و تحلیل متن خود و ارزیابی آن می‌توانند نظرات و پیشنهادات سازنده‌ی خود را برای مجری ارسال دارند. ضمناً، در صورت وجود هر نوع ابهام یا پرسش و نیاز به راهنمایی می‌توانند 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 سامانه ثبت می‌شود و بسته به مورد پاسخ مربوطه برای کاربر ارسال می‌گردد.

امروزه بسیاری از امور پژوهشی زمان بر - که زمانی صرفاً توسط انسان انجام می‌شد- به ماشین سپرده شده که این امر خود به روند انجام پژوهش و امور در جوامع انسانی شتاب بیشتری داده است. هدف از پژوهش حاضر آن بوده تا جهت استخراج و تحلیل خودکار حروف اضافه و ربط فارسی نرم افزاری تهیه شود. نرم افزاری که بتواند به محققان حوزه ی زبان در نگارش مقالات علمی و تحلیل آماری مقوله های دستوری یاری برساند.

ایده ی اولیه ی طراحی چنین نرم افزاری به سال 1386 باز می‌گردد. در آن سال با مجموعه ای از نرم افزارها که توسط پروفسور تام کاب طراحی و ارائه گردیده بود آشنا شدم. این سال، سال آغازین دوره ی دکتری اینجانب در حوزه ی زبان شناسی رایانه ای نیز بود. یکی از نرم افزارهایی که توجه ام را بسیار به خود جلب کرد نرم افزاری بود که واژگان موجود در متون انگلیسی را از نظر فراوانی رخداد مورد بررسی قرار می داد. آن نرم افزار را برای فارسی امتحان کردم که حاصل آن مقاله ای بود که در سال 2008 با عنوان:

Can an English word frequency counter meet the requirements of Persian automatic indexing?  
International Journal of Theoretical Linguistics, Vol.2, No.1, PP.18-29 به چاپ رسید.

در آن پژوهش دریافتم که نرم افزارهای تحلیل آماری به خاطر ترکیبات خاص واژگانی و قواعد واژه سازی همچون در

http://www.samaneh.com/2711/test/Step1.aspx

### تصویر ۳: نمایش مرحله‌ی اول تحلیل متن در نرم افزار

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



تصویر ۴: نمایش مرحله‌ی دوم تحلیل متن در نرم افزار

پس از کلید 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 فایل‌ها را که در رایانه‌ی شخصی خود به صورت یک فایل doc یا docx و یا txt ذخیره کرده، انتخاب نماید. در صورتی که کاربر گزینه‌ی "دریافت متن از طریق ادیتور" را برگزیند می‌تواند بخش یا تمام اطلاعات موجود در یک فایل دارای پسوند doc یا docx و یا txt را از روی رایانه‌ی شخصی کپی و در جعبه‌ی ادیتور وارد نماید. پس از انتخاب یکی از دو حالت فوق بر روی دکمه‌ی "مرحله بعد" کلید زده می‌شود.



تصویر ۵: نمایش مرحله‌ی دوم تحلیل متن در نرم افزار (ادامه تصویر ۴)

پس از انتخاب متن مورد پردازش (انتخاب فایل و یا کپی بخشی از فایل و ثبت آن در جعبه‌ی ادیتور)، لازم است نحوه‌ی مرتب سازی نتایج مشخص گردد. برای مرتب سازی نتایج، سه انتخاب وجود دارد که عبارتند از: الفبایی، فرکانس صعودی و فرکانس نزولی. در مرتب سازی الفبایی، نتایج بدون توجه به فراوانی رخداد و صرفاً براساس ترتیب الفبایی ردیف می شوند. در فرکانس صعودی مدخل‌ها براساس فراوانی از کمترین به بزرگترین ردیف می شوند. برعکس، در فرکانس نزولی واژه‌ها براساس فراوانی رخداد از فراوان ترین به نادرترین واژه ردیف می شوند. پس از انتخاب نوع مرتب سازی، روی دکمه "مرحله بعد" کلیک زده می شود.



تصویر ۶: نمایش مرحله‌ی سوم تحلیل متن در نرم افزار

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 (تصویر ۶) و از درایو مربوطه و موجود در رایانه‌ی شخصی فایل مورد نظر انتخاب می‌گردد (تصویر ۷).

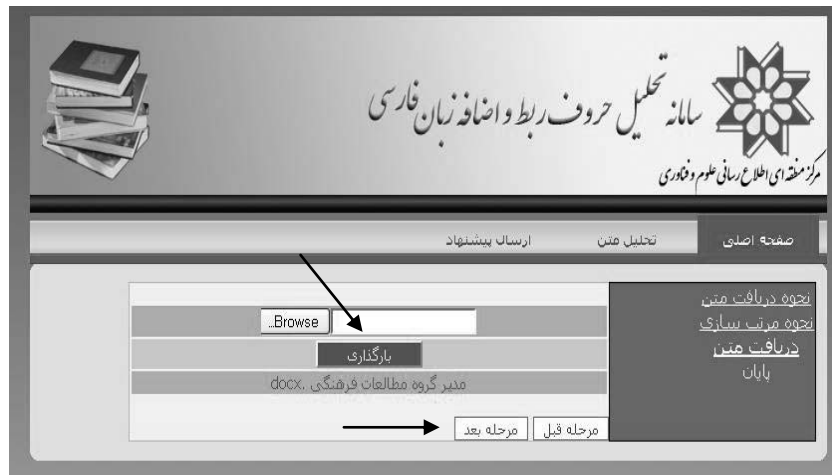


تصویر ۷-۱. نمایش مرحله‌ی چهارم تحلیل متن در نرم افزار (جستجوی فایل)



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

پس از انتخاب فایل از درایو مربوطه، مسیر فایل بارگذاری شده در جعبه‌ی روبروی گزینه‌ی "Browse" نشان داده می‌شود. به عنوان مثال، در تصویر ۷-۱ فایل "مدیر گروه مطالعات فرهنگی.docx" جستجو شد و در تصویر ۷-۲ مسیر فایل مربوطه جلوی دکمه‌ی "Browse" درج گردید. سپس، بر روی دکمه‌ی "بارگذاری" کلیک زده می‌شود.



تصویر ۸: نمایش مرحله‌ی پنجم تحلیل متن در نرم افزار

پس از کلیک زدن بر روی دکمه‌ی "بارگذاری"، در تصویر ۷-۲، فایل بارگذاری شده به صورتی که در تصویر ۸ مشاهده می‌شود، نمایش داده می‌شود. به عنوان مثال، فایل بارگذاری شده در تصویر ۸، فایل "مدیر گروه مطالعات فرهنگی. docx" است که در ذیل دکمه‌ی "بارگذاری" درج گردیده است. پس از انجام بارگذاری، روی دکمه‌ی "مرحله بعد" کلیک زده می‌شود.



تصویر ۹: نمایش مرحله‌ی ششم تحلیل متن در نرم افزار

در این مرحله، نرم افزار آماده‌ی انجام تحلیل متن است. برای این منظور تنها کافی است بر روی دکمه‌ی "تحلیل متن" کلیک زده شود. چنانچه کاربر از ادامه‌ی کار منصرف شده باشد، می‌تواند دکمه‌ی "انصراف" را انتخاب نموده، از برنامه خارج شود.

رف	واژه	فراوانی مطلق	فراوانی نسبی	درصد فراوانی نسبی	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	یا	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	برای	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
13	م	12	0.002	0.2000	1335	21.5432

تصویر ۱۰-۱: نمایش مرحله‌ی هفتم تحلیل متن در نرم افزار (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 نرم افزار به دو صورت است: ۱- به صورت کلی (تصویر ۱۰-۱) و ۲- به صورت مجزا (تصویر ۱۰-۲). در واقع، می‌توان اطلاعات آماری مربوط به واژگان موجود در گروه‌های شش گانه را به صورت جدا

محاسبه نمود و یا همه‌ی آن‌ها را در قالب یک جدول و فهرست واحد، نشان داد. پیش فرض در نرم افزار، "نمایش نتایج به صورت مجزا" تعیین گردیده که البته کاربر در صورت تمایل می‌تواند "نمایش نتایج به صورت کلی" را برگزیند. نرم افزار 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" کلید زد (تصویر ۱۰-۲).

شرف	واژه	فراوانی مطلق	فراوانی نسبی	درصد فراوانی نسبی	فراوانی تجمعی	درصد فراوانی تجمعی
۱	هم چون	1	0.0002	0.0161	1471	23.7488
۲	هم جو	0	0.0000	0.0000	1472	23.7649
۳	مگر از	0	0.0000	0.0000	1472	23.7649
۴	جز که	0	0.0000	0.0000	1472	23.7649
۵	جز از	0	0.0000	0.0000	1472	23.7649
۶	به جز	0	0.0000	0.0000	1472	23.7649
۷	چون که	0	0.0000	0.0000	1472	23.7649
۸	به جز از	0	0.0000	0.0000	1472	23.7649
۹	هم چون که	0	0.0000	0.0000	1472	23.7649
۱۰	به جز که	0	0.0000	0.0000	1472	23.7649
۱۱	مگر که	0	0.0000	0.0000	1472	23.7649
۱۲	از برای	0	0.0000	0.0000	1472	23.7649
	مجموع	1	0.0002 = (1/6194)	0.0161	1472	23.7649

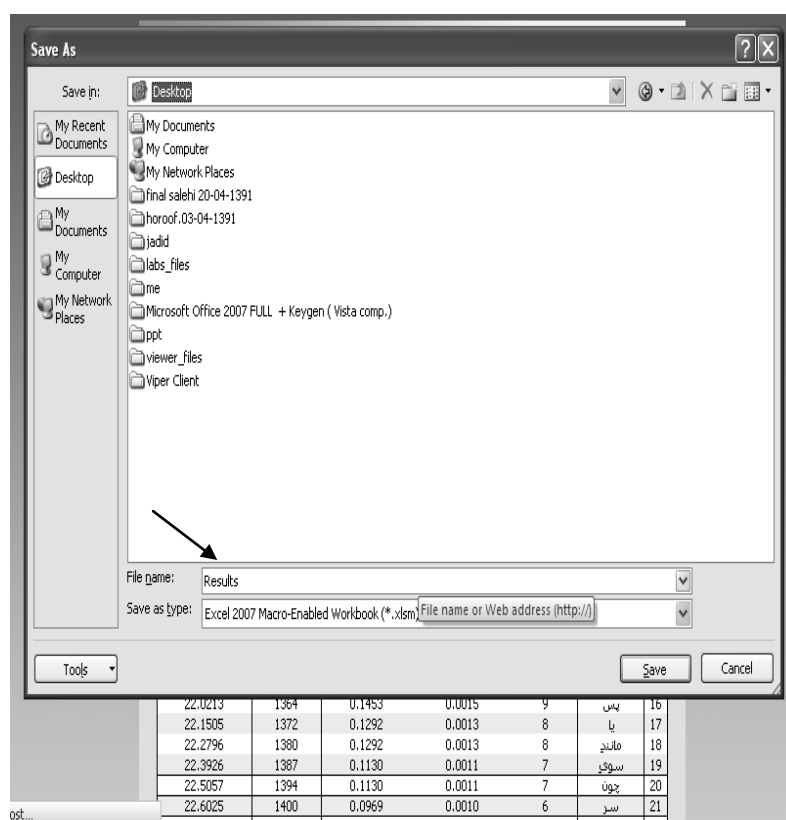
  

شرف	واژه	فراوانی مطلق	فراوانی نسبی	درصد فراوانی نسبی	فراوانی تجمعی	درصد فراوانی تجمعی
۱	در	199	0.0321	3.2128	598	9.6545
۲	به	180	0.0291	2.9060	778	12.5605
۳	از	165	0.0266	2.6639	943	15.2244
۴	را	125	0.0202	2.0181	1068	17.2425
۵	با	74	0.0119	1.1947	1235	19.9387
۶	بر	25	0.0040	0.4036	1260	20.3423

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

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





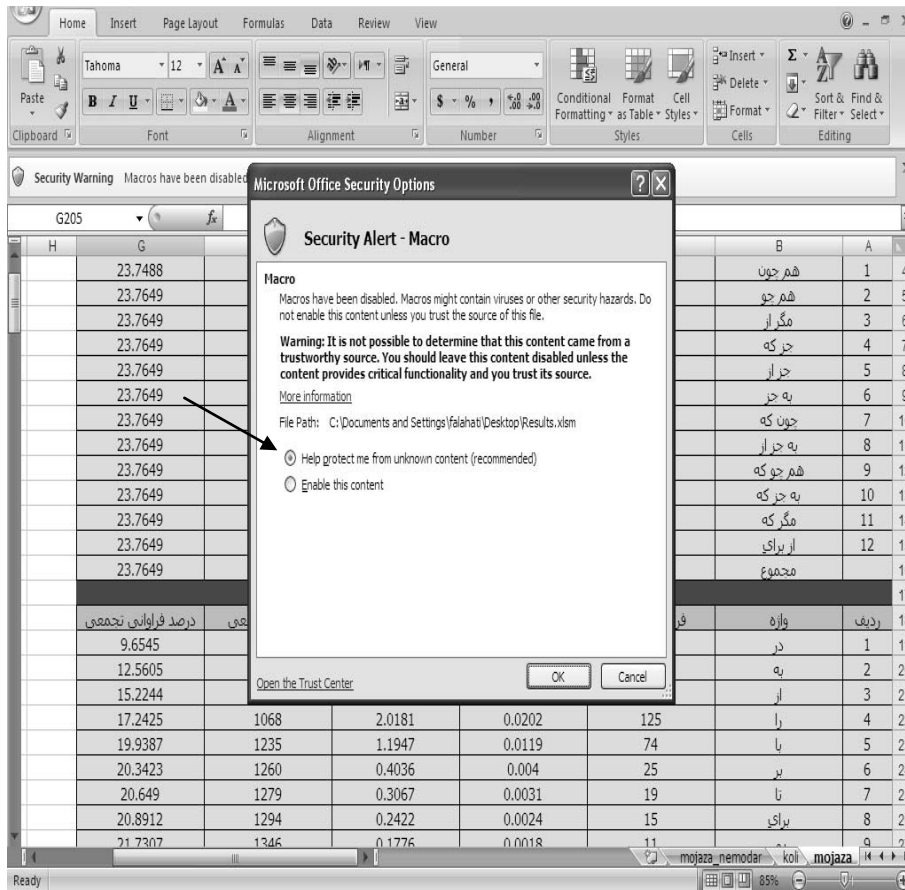
تصویر ۱۱: نمایش مرحله‌ی هشتم تحلیل متن در نرم افزار

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 پسوندی مانند ۱، ۲ و ... را اضافه نمود مانند "Results1"، "Results2" و نظیر آن. فایل حاصله را می‌توان با پسوند xism در درایو مورد نظر در رایانه‌ی شخصی ذخیره نمود.

ردیف	واژه	فراوانی مطلق	فراوانی نسبی	درصد فراوانی نسبی	فراوانی تجمعی	درصد فراوانی تجمعی
1	در	199	0.0321	3.2128	598	9.6545
2	به	180	0.0291	2.906	778	12.5605
3	از	165	0.0266	2.6639	943	15.2244
4	را	125	0.0202	2.0181	1068	17.2425
5	با	74	0.0119	1.1947	1235	19.9387
6	بر	25	0.004	0.4036	1260	20.3423
7	تا	19	0.0031	0.3067	1279	20.649
8	برای	15	0.0024	0.2422	1294	20.8912
9		11	0.0018	0.1776	1346	21.7307
10	مجموع	1	$(1/6194) = 0.0002$	0.0161	1472	23.7649

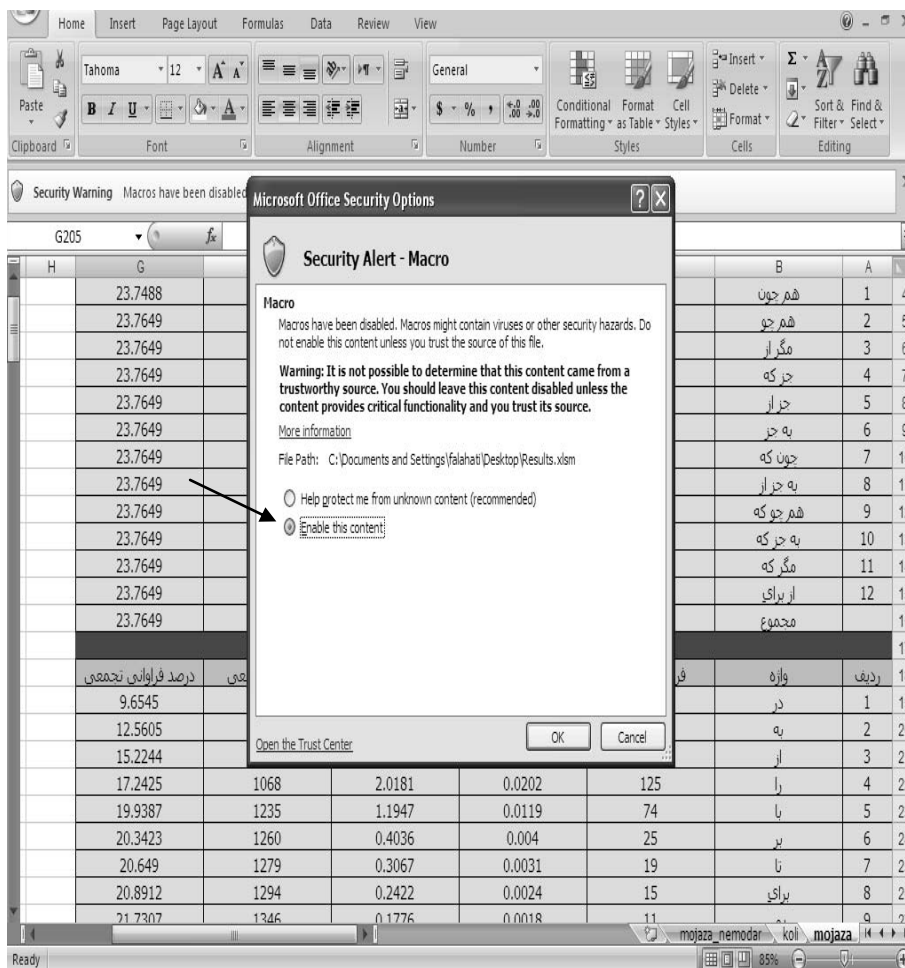
تصویر ۱۲: نمایش مرحله‌ی نهم تحلیل متن در نرم افزار

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 درهم کرد اطلاعات مربوط به گروه‌های شش‌گانه را دریافت نماید، در تصویر ۱۲ به جای شش جدول، یک جدول واحد که حاوی کل واژگان شش‌گروه باشد، نمایش داده خواهد شد.



تصویر ۱۳-۱: نمایش مرحله‌ی دهم تحلیل متن در نرم افزار

در بالای جدولی که در تصویر ۱۲ ارایه گردیده، گزینه‌ای تحت عنوان "Options" وجود دارد. برای اینکه نرم افزار 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 می شود. پیش فرض در این گزینه، "Help protect me from unknown content" می باشد. لازم است این پیش فرض به "Enable this content" تغییر یابد.



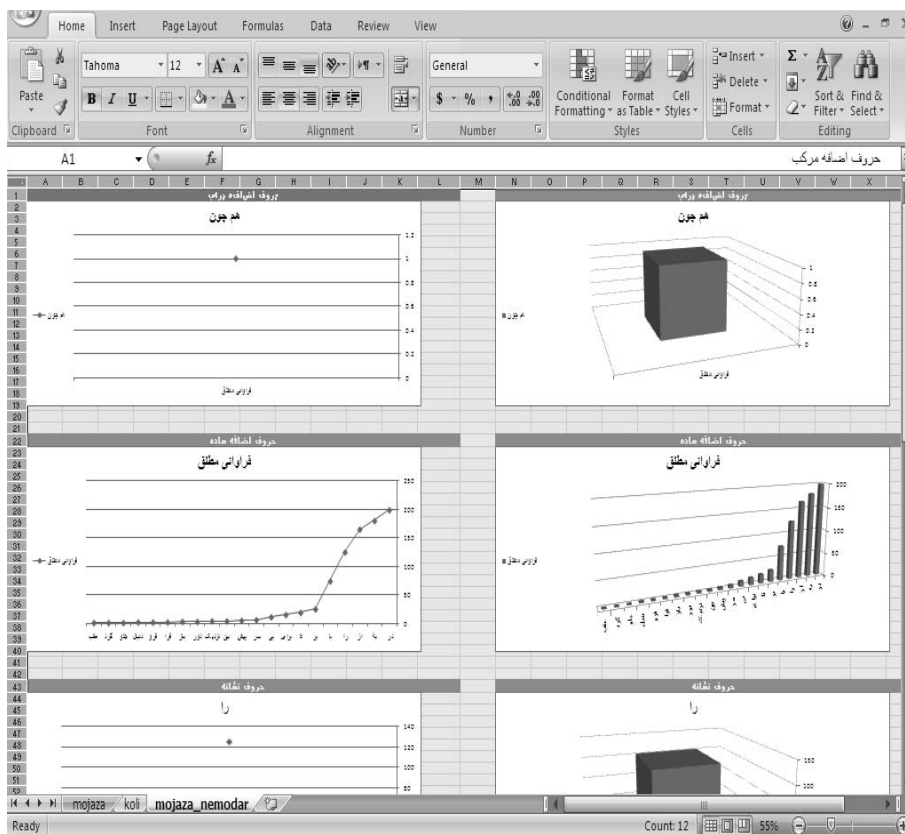
تصویر ۱۳-۲: نمایش مرحله‌ی دهم تحلیل متن در نرم افزار (ادامه تصویر ۱۳-۱)

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

H	G	F	E	D	C	B	A
	23.7649	1472	0	0	0	به نزدیکی	30
	23.7649	1472	0	0	0	از واسطه‌ی	31
	23.7649	1472	0	0	0	اندر جنب	32
	23.7649	1472	0	0	0	بر کردار	33
	23.7649	1472	0	0	0	به کردار	34
	23.7649	1472	0	0	0	از قبل	35
	23.7649	1472	0	0	0	از بهر	36
	23.7649	1472	0	0	0	به جای	37
	23.7649	1472	0	0	0	گذشت از	38
	23.7649	1472	0	0	0	بیرون	39
	23.7649	1472	0	0	0	عبر از	40
	23.7649	1472	0	0	0	به سوی	41
	23.7649	1472	0	0	0	به جهت	42
	23.7649	1472	0	0	0	از راه	43
	23.7649	1472	0	0	0	در باب	44
	23.7649	1472	0	0	0	بر جای	45
	23.7649	1472	0	0	0	به غیر	46
	23.7649	1472	0,5489	(34/6194) = 0,0055	34	مجموع	201
							202
							203
							204
							205
							206
							207
							208
							209

تصویر ۱۴: نمایش مرحله‌ی یازدهم تحلیل متن در نرم افزار

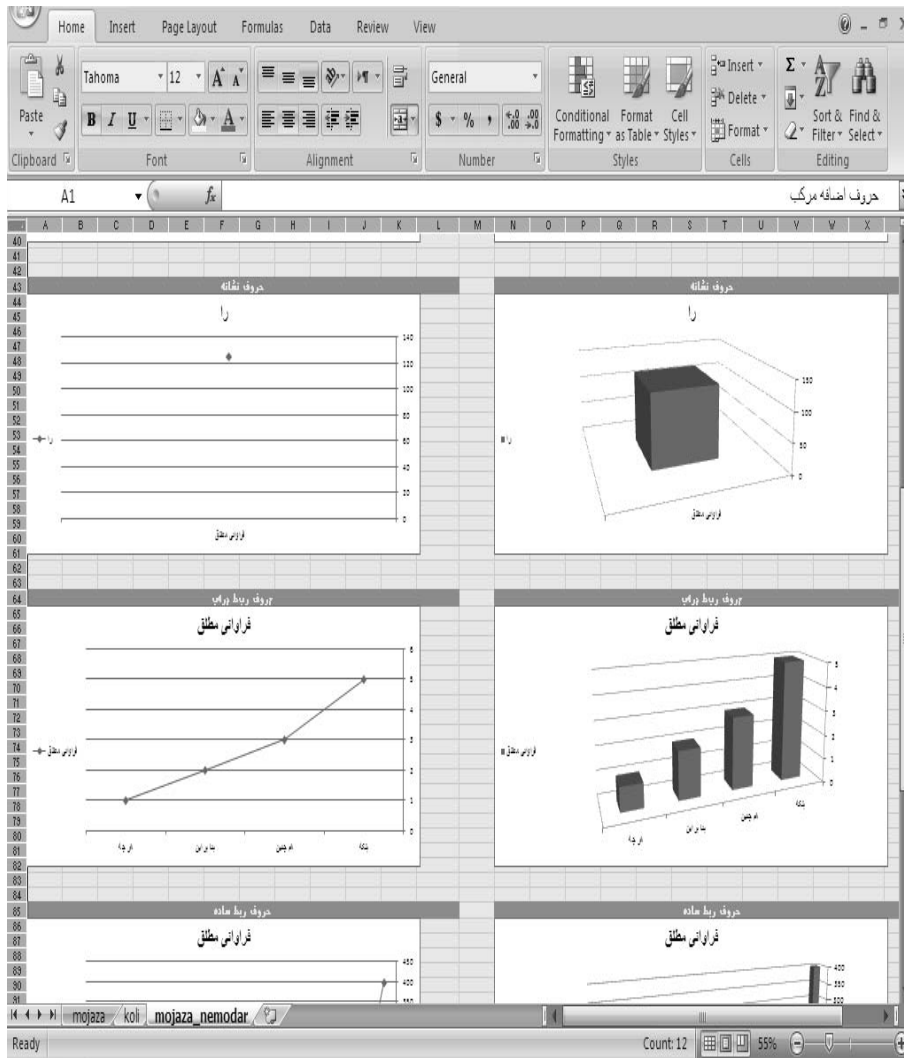
در این مرحله، چنانچه کاربر بخواهد از اطلاعات موجود در جداول استفاده نماید می‌تواند آن‌ها را در فایل Word ذخیره نماید و در مقاله‌ی خود مورد استفاده قرار دهد. حال اگر مایل باشد از داده‌های خود نموداری نیز تهیه کند، می‌تواند بر روی گزینه‌ی "رسم نمودار" در تصویر ۱۴ کلیک بزند.



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

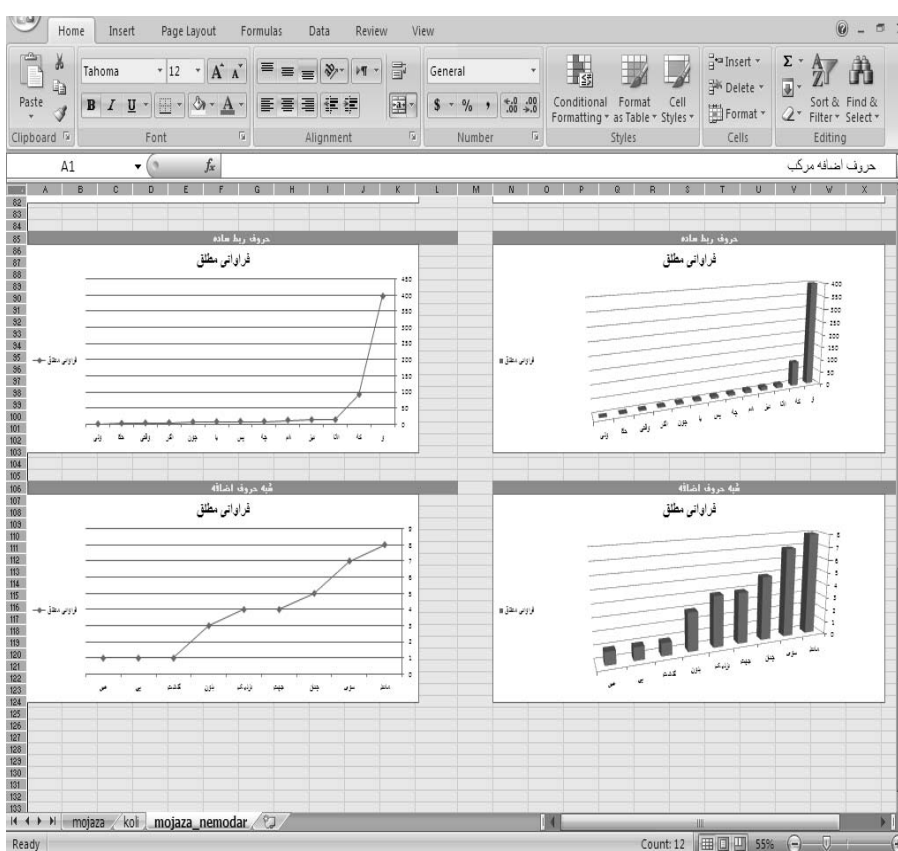
پس از کلید زدن بر روی گزینه‌ی "رسم نمودار" در تصویر ۱۴، برای هر یک از گروه‌های شش‌گانه، دو نوع نمودار (یک نمودار ستونی و یک نمودار خطی) به شرح تصویرهای ۱۵-۱ الی ۱۵-۳ ترسیم می‌گردد.

تصویر ۱۵-۱ نمودارهای مربوط به حروف اضافه‌ی مرکب و حروف اضافه‌ی ساده را نشان می‌دهد.



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

تصویر ۱۵-۲ نمودارهای مربوط به حرف نشانه "را" و حروف ربط مرکب را نشان می دهد.

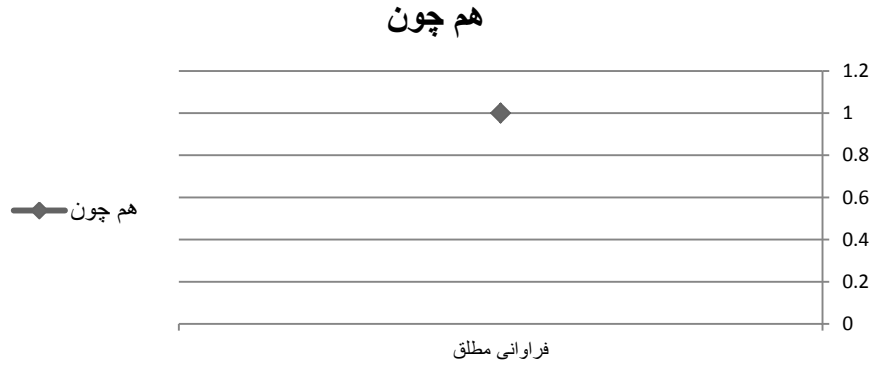


تصویر ۱۵-۳: نمایش مرحله‌ی دوازدهم تحلیل متن در نرم افزار (حروف ربط ساده و شبه حروف اضافه)

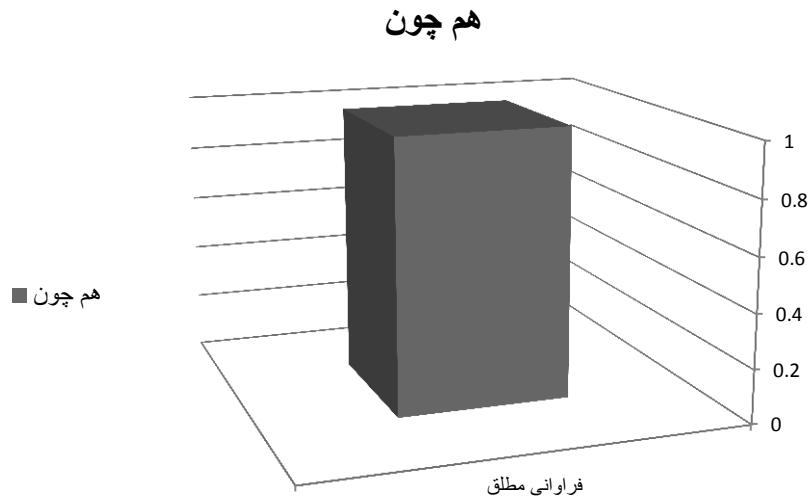
تصویر ۱۵-۳ نمودارهای مربوط به حروف ربط ساده و شبه حروف اضافه را به تصویر می کشد.

چنانچه کاربر مایل باشد از یک نمودار خاص با وضوح بیشتر استفاده نماید، می تواند نمودار مربوطه را انتخاب و در فایل Word خود کپی نماید. در این صورت می تواند ابعاد نمودار و اندازه‌ی قلم و مختصات دیگر را متناسب با نیاز خود تغییر دهد. در این صورت، اطلاعات موجود در تصاویر ۱۵-۱ الی ۱۵-۳ به شکل تصاویر ۱۶-۱ الی ۱۶-۱۲ تغییر خواهد کرد.



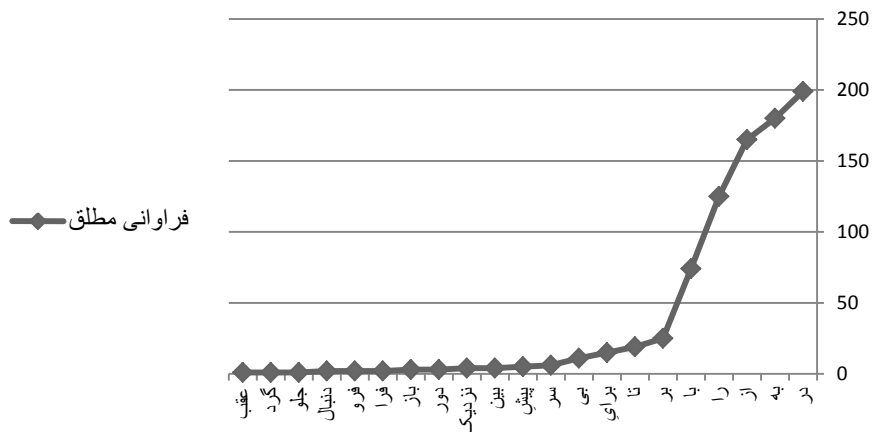


تصویر ۱۶-۱: نمایش اطلاعات مربوط به حروف اضافه‌ی مرکب (نمودار خطی)



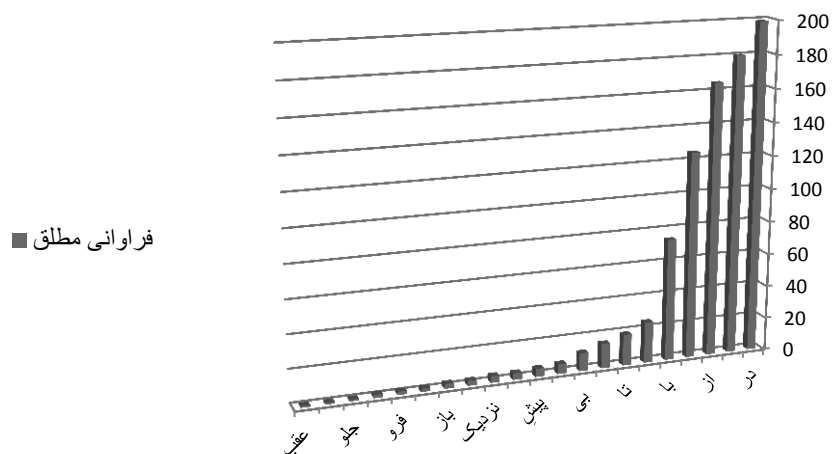
تصویر ۱۶-۲: نمایش اطلاعات مربوط به حروف اضافه‌ی مرکب (نمودار ستونی)

### فراوانی مطلق



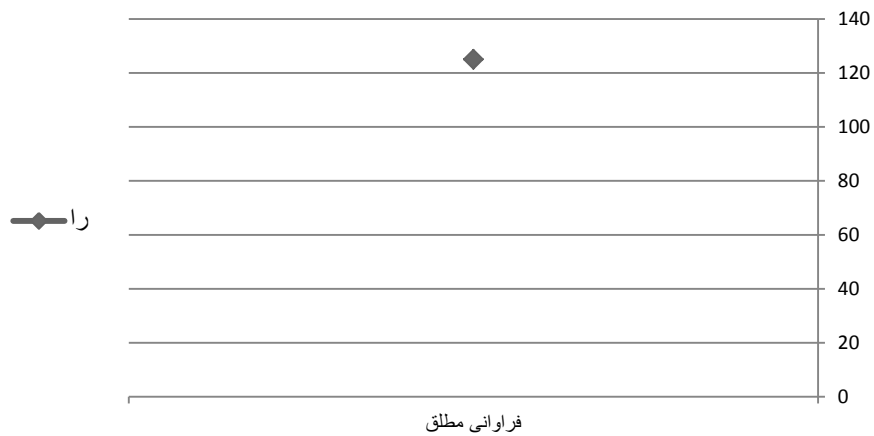
تصویر ۱۶-۳: نمایش اطلاعات مربوط به حروف اضافه‌ی ساده (نمودار خطی).

### فراوانی مطلق



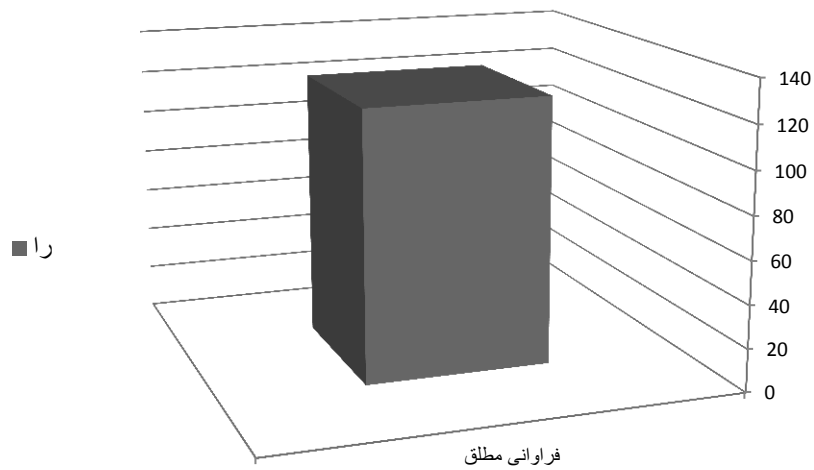
تصویر ۱۶-۴: نمایش اطلاعات مربوط به حروف اضافه‌ی ساده (نمودار ستونی).

را



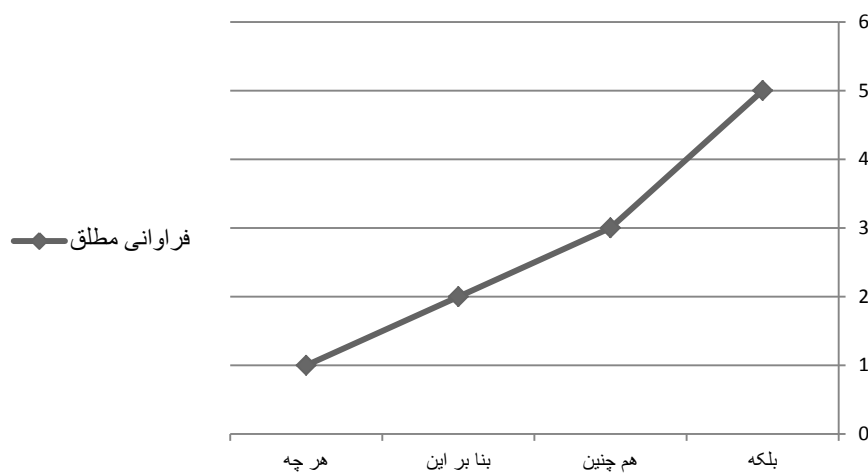
تصویر ۱۶-۵: نمایش اطلاعات مربوط به حرف نشانه "را" (نمودار خطی).

را



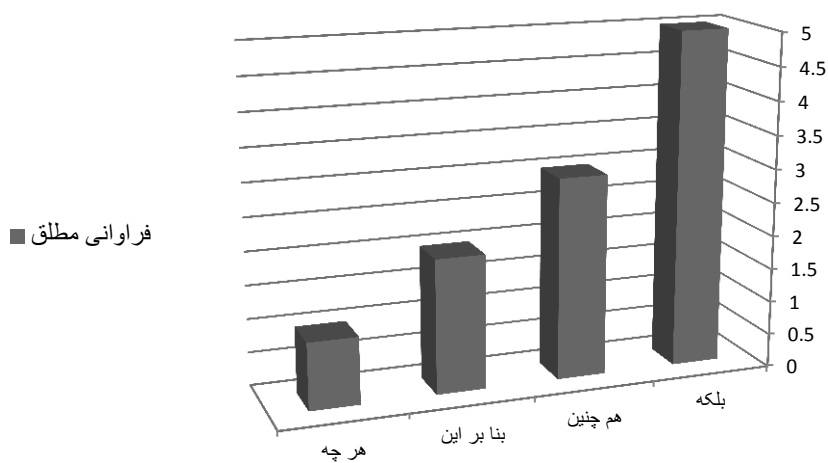
تصویر ۱۶-۶: نمایش اطلاعات مربوط به حرف نشانه "را" (نمودار ستونی).

### فراوانی مطلق



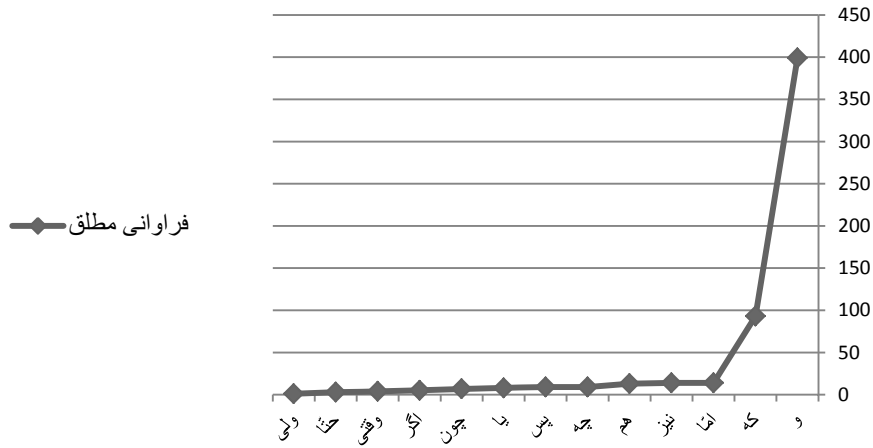
تصویر ۱۶-۷: نمایش اطلاعات مربوط به حروف ربط مرکب (نمودار خطی).

### فراوانی مطلق



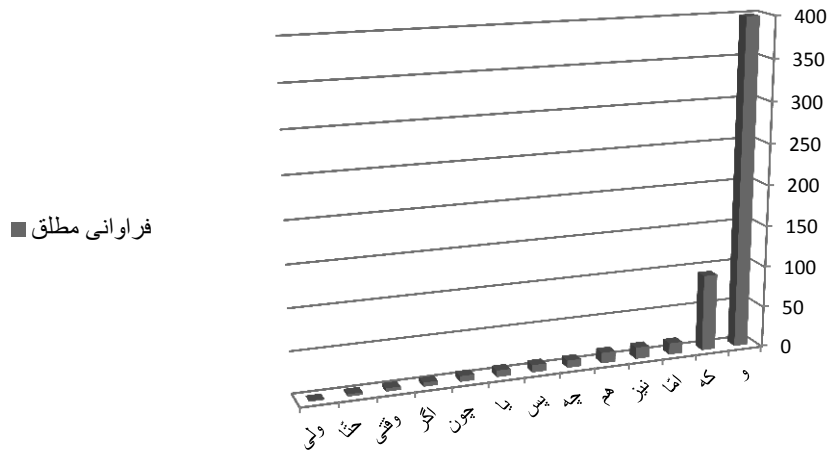
تصویر ۱۶-۸: نمایش اطلاعات مربوط به حروف ربط مرکب (نمودار ستونی).

فراوانی مطلق



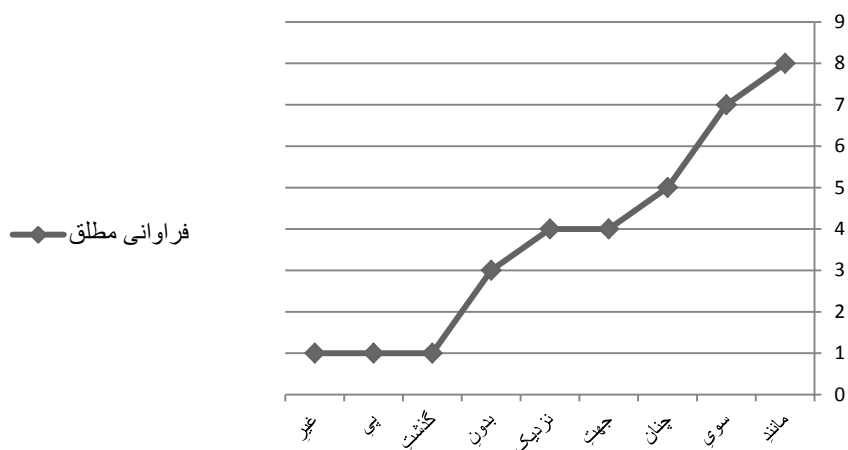
تصویر ۱۶-۹: نمایش اطلاعات مربوط به حروف ربط ساده (نمودار خطی).

فراوانی مطلق



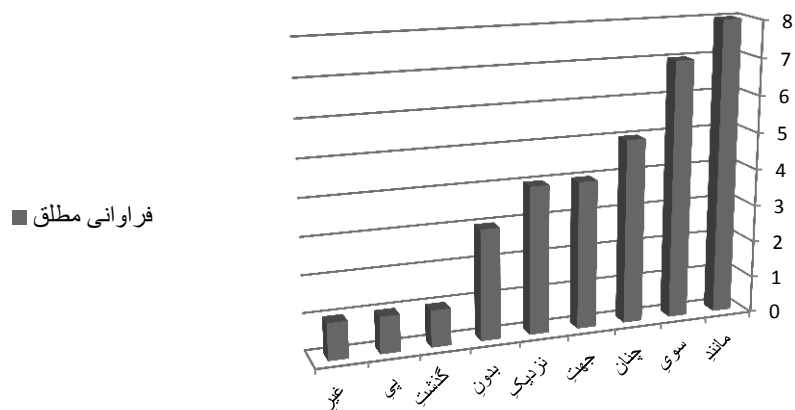
تصویر ۱۶-۱۰: نمایش اطلاعات مربوط به حروف ربط ساده (نمودار ستونی).

## فراوانی مطلق



تصویر ۱۱-۱۶: نمایش اطلاعات مربوط به شبه حروف اضافه (نمودار خطی).

## فراوانی مطلق



تصویر ۱۲-۱۶: نمایش اطلاعات مربوط به شبه حروف اضافه (نمودار ستونی).

## زمینه‌هایی برای مطالعات بیشتر:

- در پژوهش حاضر صرفاً بر روی مقوله‌های بسته‌ای چون حروف (حروف اضافه، نشانه و ربط) کار شد. می‌توان نرم افزار مربوطه را در قالب پژوهشی دیگر به سایر حوزه‌ها و مقوله‌های دستوری گسترش داد.
- به دلیل پیچیدگی فرایند 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 نرم افزار این دو حیطة به شش حیطة افزایش پیدا نمود.
- از طریق این نرم افزار و با تحلیل پیکره‌ی زبانی مناسب می‌توان حروف فعال و غیر فعال (مرده) را در زبان فارسی مشخص نمود. منظور از حروف فعال حرفی است که همچنان در فارسی امروز کاربرد دارد و مراد از حروف غیر فعال حرفی است که در فارسی امروز به کار برده نمی‌شوند.

**خلاصه:**

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 ساده 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 املائی، برای هر واژه، تمامی صورت‌های املائی احتمالی، 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 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.



## فهرست منابع:

- ۱- احمدی گیوی، ح. و انوری، ح. (۱۳۸۵). دستور زبان فارسی ۱ (ویرایش سوم). تهران: انتشارات فاطمی.
- ۲- باطنی، م. ر. (۱۳۷۲). توصیف ساختمان دستوری زبان فارسی (چاپ پنجم). تهران: انتشارات امیرکبیر.
- ۳- شریعت، م. ج. (۱۳۸۴). دستور زبان فارسی (چاپ هشتم). تهران: انتشارات اساطیر.
- ۴- فرشید ورد، خ. (۱۳۸۲). دستور مفصل امروز (چاپ اول). تهران: انتشارات سخن.
- ۵- لازار، ژ. (۱۳۸۴). دستور زبان فارسی معاصر (ترجمه مهستی بحرینی). تهران: هرمس، مرکز بین‌المللی گفتگوی تمدن‌ها. (اثر اصلی در سال ۱۹۵۷ به چاپ رسید).

## پیوست ۱:

## زنجیره املائی حروف اضافه ساده و مرکب جهت رفع مشکل تنوع املائی

حروف اضافه ساده		حروف اضافه مرکب	
از#۱	سواى#۲۶	به جز#۴۸	جز که#۵۳
الا#۲	پایین#۲۷	بهجز#۴۸	مگر که#۵۴
آلا#۲	پائین#۲۷	بجز#۴۸	مگر که#۵۴
اندر#۳	پشت#۲۸	بهجز از#۴۹	مگر که#۵۴
ایدون#۴	نزد#۲۹	به جز از#۴۹	هم چون#۵۵
به#۵	جلو#۳۰	به جز از#۴۹	همچون#۵۵
با#۶	عقب#۳۱	به جز از#۴۹	همچون#۵۵
باز#۷	پهلوی#۳۲	بهجز از#۴۹	هم چو#۵۶
بر#۸	اندرون#۳۳	بجز از#۴۹	همچو#۵۶
بی#۹	درون#۳۴	بجز از#۴۹	همچو#۵۶
تا#۱۰	بالا#۳۵	بجز از#۴۹	همچو که#۵۷
جز#۱۱	دم#۳۶	جز از#۵۰	هم چو که#۵۷
چو#۱۲	دنبال#۳۷	جز از#۵۰	همچو که#۵۷
در#۱۴	زیر#۳۸	جز از#۵۰	هم چو که#۵۷
را#۱۵	سر#۳۹	مگر از#۵۱	هم چو که#۵۷
زی#۱۶	کنار#۴۰	مگر از#۵۱	همچو که#۵۷
فا#۱۷	گرد#۴۱	مگر از#۵۱	همچو که#۵۷
فرا#۱۸	دور#۴۲	به جز که#۵۲	همچو که#۵۷
فرو#۱۹	میان#۴۳	به جز که#۵۲	از برای#۵۸
20#	بین#۴۴	بهجز که#۵۲	از برای#۵۸
مگر#۲۱	نزدیک#۴۵	بجز که#۵۲	از برای#۵۸
وا#۲۲	پیش#۴۶	به جز که#۵۲	از برای#۵۸
که#۲۳	پیش#۴۶	بجز که#۵۲	از برای#۵۸
زیر#۲۴	برای#۴۷	بجز که#۵۲	از برای#۵۸
الی#۲۵	برای#۴۷	بهجز که#۵۲	چونکه#۵۹
		جز که#۵۳	چون که#۵۹
		جز که#۵۳	چونکه#۵۹

پیوست ۲:

## زنجیره املائی شبه حروف اضافه جهت رفع مشکل تنوع املائی

شبه حروف اضافه						
برکردار#۱۰۳	بهرکردار#۹۷	بجای#۹۱	ازسر#۸۷	ازپی#۸۳	بیرونناز#۷۲	بهجهت#۶۰
درباب#۱۰۴	بکردار#۹۷	بدون#۹۲	ازقبل#۸۸	ازپی#۸۳	بیرونز#۷۲	بجهت#۶۰
درباب#۱۰۴	به نزدیک#۹۸	بدون#۹۲	ازقبل#۸۸	ازپی#۸۴	بیرونز#۷۲	بهجهت#۶۰
در باب#۱۰۴	به نزدیک#۹۸	بسان#۹۳	از قبل#۸۸	از پی#۸۵	جدا از#۷۳	بهجهت#۶۰
در باب#۱۰۴	به نزدیک#۹۸	بسان#۹۳	از قبل#۸۸	از پی#۸۶	جدا از#۷۳	بهجهت#۶۰
درباب#۱۰۴	بندیدک#۹۸	به سر#۹۴	از قبل#۸۸	از پی#۸۷	جدا از#۷۳	بجهت#۶۰
درباب#۱۰۴	به نزدیک#۹۸	به سر#۹۴	از قبل#۸۸	از پی#۸۸	چنان#۷۴	بهر#۶۱
در جنب#۱۰۵	بندیدک#۹۸	بهسر#۹۴	از واسطه‌ی#۸۹	از پی#۸۹	چونان#۷۵	بهر#۶۱
در جنب#۱۰۵	باسر#۹۹	بهسر#۹۴	از واسطه‌ی#۸۹	از پی#۹۰	غیر از#۷۶	بیرون#۶۲
در جنب#۱۰۵	باسر#۹۹	بسر#۹۴	از واسطه‌ی#۸۹	از پی#۹۱	غیر از#۷۶	بیرون#۶۲
در جنب#۱۰۵	برجای#۱۰۰	بسر#۹۴	از واسطه‌ی#۸۹	از پی#۹۲	غیر از#۷۶	پی#۶۳
در جنب#۱۰۵	برجای#۱۰۰	بسوی#۹۵	از واسطه‌ی#۸۹	از پی#۹۳	گذشته از#۷۷	پی#۶۳
در جنب#۱۰۵	بر جای#۱۰۰	به سوی#۹۵	از واسطه‌ی#۸۹	از پی#۹۴	گذشته از#۷۷	جهت#۶۵
در حق#۱۰۶	برجای#۱۰۰	به سوی#۹۵	از واسطه‌ی#۸۹	از پی#۹۵	گذشت از#۷۸	جهت#۶۵
در حق#۱۰۶	بر جای#۱۰۰	بهسوی#۹۵	از واسطه‌ی#۸۹	از پی#۹۶	گذشته از#۷۸	دون#۶۶
در حق#۱۰۶	برجای#۱۰۰	بسوی#۹۵	از واسطه‌ی#۸۹	از پی#۹۷	مثل#۷۹	سوی#۶۷
در حق#۱۰۶	برسان#۱۰۱	بهسوی#۹۵	از واسطه‌ی#۸۹	از پی#۹۸	مثل#۷۹	سوی#۶۷
در حق#۱۰۶	برسان#۱۰۱	به غیر#۹۶	اندر جنب#۹۰	از پی#۹۹	غیر#۸۰	گذشت#۶۸
در حق#۱۰۶	برسر#۱۰۲	به غیر#۹۶	اندر جنب#۹۰	از پی#۱۰۰	غیر#۸۰	گذشت#۶۸
چنان چون#۱۰۷	برسر#۱۰۲	بغیر#۹۶	اندر جنب#۹۰	از پی#۱۰۱	از بهر#۸۲	مانند#۶۹
چنانچون#۱۰۷	برسر#۱۰۲	بهغیر#۹۶	اندر جنب#۹۰	از پی#۱۰۲	از بهر#۸۲	مانند#۶۹
چنانچون#۱۰۷	بر سر#۱۰۲	بغیر#۹۶	اندر جنب#۹۰	از پی#۱۰۳	از بهر#۸۲	ماننده#۷۰
بر کردار#۱۰۳	بر کردار#۱۰۳	بهغیر#۹۶	اندر جنب#۹۰	از پی#۱۰۴	از بهر#۸۲	نزدیک#۷۱
بر کردار#۱۰۳	برکردار#۹۷	به کردار#۹۷	به جای#۹۱	از پی#۱۰۵	از بهر#۸۲	نزدیک#۷۱
بر کردار#۱۰۳	برکردار#۹۷	به کردار#۹۷	به جای#۹۱	از پی#۱۰۶	از بهر#۸۲	بیرون از#۷۲
بر کردار#۱۰۳	برکردار#۹۷	به کردار#۹۷	بجای#۹۱	از پی#۱۰۷	از پی#۸۳	بیرون ز#۷۲
بر کردار#۱۰۳	بکردار#۹۷	بهجای#۹۱	ازسر#۸۷	از پی#۸۳	از پی#۸۳	بیرون ز#۷۲

پیوست ۳:

زنجیره املائی حرف نشانه و حروف ربط ساده جهت رفع مشکل تنوع املائی

حرف نشانه	حروف ربط ساده
را#۱۰۸	و#۱۰۹
	یا#۱۱۰
	ولی#۱۱۱
	اما#۱۱۲
	افا#۱۱۲
	پس#۱۱۴
	سپس#۱۱۵
	هم#۱۱۷
	نیز#۱۱۸
	لیک#۱۱۹
	ولیکن#۱۱۹
	لیکن#۱۱۹
	لکن#۱۱۹
	چه#۱۲۰
	حتّا#۱۲۱
	حتّی#۱۲۱
	حتی#۱۲۱
	حتا#۱۲۱
	چون#۱۲۲
	که#۱۲۳
	خواه#۱۲۴
	کجا#۱۲۵
	وقتی#۱۲۶
	اگر#۱۲۷
	زیرا#۱۲۸

پیوست ۴:

زنجیره املائی حروف ربط مرکب جهت رفع مشکل تنوع املائی

حروف ربط مرکب					
هر قدر#۱۵۱	بعلتاین که#۱۴۹	مگر اینکه#۱۴۷	تا اینکه#۱۴۳	در نتیجه#۱۳۶	بلکه#۱۳۰
هر قدر#۱۵۱	به‌علتاین که#۱۴۹	مگر اینکه#۱۴۷	تا اینکه#۱۴۳	در نتیجه#۱۳۶	با این همه#۱۳۱
هر کجا#۱۵۲	به علتاین که#۱۴۹	الا اینکه#۱۴۸	تا اینکه#۱۴۳	بنابراین#۱۳۷	باین همه#۱۳۱
هر کجا#۱۵۲	به علت این که#۱۴۹	الا اینکه#۱۴۸	تا اینکه#۱۴۳	بنابراین#۱۳۷	باین همه#۱۳۱
هر کجا#۱۵۲	به‌علتاین که#۱۴۹	الا اینکه#۱۴۸	تا این که#۱۴۳	بنا بر این#۱۳۷	باین همه#۱۳۱
چندان که#۱۵۳	به علتاین که#۱۴۹	آلا اینکه#۱۴۸	تا این که#۱۴۳	بنابر این#۱۳۷	باین همه#۱۳۱
چندانکه#۱۵۳	به‌علت این که#۱۴۹	آلا اینکه#۱۴۸	تا این که#۱۴۳	بنابر این#۱۳۷	با این همه#۱۳۱
چندانکه#۱۵۳	به علت این که#۱۴۹	آلا اینکه#۱۴۸	اگر چه#۱۴۴	بنا بر این#۱۳۷	با این همه#۱۳۱
هر چند#۱۵۴	به علت این که#۱۴۹	الا اینکه#۱۴۸	اگر چه#۱۴۴	و گر نه#۱۳۸	هم چنین#۱۳۲
هر چند#۱۵۴	به‌علت این که#۱۴۹	الا اینکه#۱۴۸	اگر چه#۱۴۴	و گر نه#۱۳۸	همچنین#۱۳۲
هر چند#۱۵۴	به‌علتاین که#۱۴۹	الا این که#۱۴۸	هر گاه#۱۴۵	و گر نه#۱۳۸	همچنین#۱۳۲
آنجا که#۱۵۵	به علتاین که#۱۴۹	الا این که#۱۴۸	هر گاه#۱۴۵	و گر نه#۱۳۸	وانگهی#۱۳۳
آنجا که#۱۵۵	به‌علتاین که#۱۴۹	الا این که#۱۴۸	هر گاه#۱۴۵	و گر نه#۱۳۸	وانگهی#۱۳۳
آنجا که#۱۵۵	به علت این که#۱۴۹	آلا این که#۱۴۸	در صورتی که#۱۴۶	و گر نه#۱۳۸	وان گهی#۱۳۳
از آنجا که#۱۵۶	به علتاین که#۱۴۹	آلا اینکه#۱۴۸	در صورتی که#۱۴۶	و گر نه#۱۳۸	معهدا#۱۳۴
از آنجا که#۱۵۶	به علت این که#۱۴۹	آلا اینکه#۱۴۸	در صورتیکه#۱۴۶	برعکس#۱۳۹	معهدا#۱۳۴
از آنجا که#۱۵۶	به علتاین که#۱۴۹	آلا این که#۱۴۸	در صورتیکه#۱۴۶	بر عکس#۱۳۹	مع هذا#۱۳۴
از آنجا که#۱۵۶	به‌علت این که#۱۴۹	آلا این که#۱۴۸	در صورتیکه#۱۴۶	برعکس#۱۳۹	با این حال#۱۳۵
از آنجا که#۱۵۶	به علت این که#۱۴۹	به‌علت این که#۱۴۹	در صورتی که#۱۴۶	و آلا#۱۴۰	با اینحال#۱۳۵
از آنجا که#۱۵۶	بعلت این که#۱۴۹	بعلتاین که#۱۴۹	در صورتیکه#۱۴۶	و الا#۱۴۰	با اینحال#۱۳۵
از آنجا که#۱۵۶	به‌علتاین که#۱۴۹	بعلتاین که#۱۴۹	مگر این که#۱۴۷	و آلا#۱۴۰	باین حال#۱۳۵
از این جهت که#۱۵۷	هر چه#۱۵۰	به‌علتاین که#۱۴۹	مگر این که#۱۴۷	و آلا#۱۴۰	باینحال#۱۳۵
از این جهت که#۱۵۷	هر چه#۱۵۰	به علتاین که#۱۴۹	مگر این که#۱۴۷	چنانکه#۱۴۱	باین حال#۱۳۵
از این جهت که#۱۵۷	هر چه#۱۵۰	بعلت این که#۱۴۹	مگر این که#۱۴۷	چنانکه#۱۴۱	باینحال#۱۳۵
از این جهت که#۱۵۷	هر قدر#۱۵۱	بعلت این که#۱۴۹	مگر این که#۱۴۷	چنان که#۱۴۱	در نتیجه#۱۳۶

پیوست ۴:

زنجیره املایی حروف ربط مرکب جهت رفع مشکل تنوع املایی (ادامه)

ادامه حروف ربط مرکب				
بمحض اینکه#۱۷۲	بشرط آن که#۱۶۹	با این که#۱۶۶	اکنون که#۱۶۲	از این جهت که#۱۵۷
بهمحض اینکه#۱۷۲	به شرط آن که#۱۶۹	با اینکه#۱۶۶	اکنون که#۱۶۲	از این جهت که#۱۵۷
بمحض اینکه#۱۷۲	بشرط آن که#۱۶۹	باین که#۱۶۶	اگر چنان چه#۱۶۳	از این جهت که#۱۵۷
بمحض اینکه#۱۷۲	بشرط آن که#۱۶۹	باین که#۱۶۶	اگر چنانچه#۱۶۳	از این جهت که#۱۵۷
بمحض این که#۱۷۲	بشرط آن که#۱۶۹	با اینکه#۱۶۶	اگر چنان چه#۱۶۳	از این جهت که#۱۵۷
به محض اینکه#۱۷۲	بشرط آن که#۱۶۹	با وجود این#۱۶۷	اگر چنان چه#۱۶۳	از این جهت که#۱۵۷
بهمحض اینکه#۱۷۲	به طوری که#۱۷۰	با وجود این#۱۶۷	اگر چنانچه#۱۶۳	از این جهت که#۱۵۷
به محض اینکه#۱۷۲	ببطوری که#۱۷۰	با وجود این#۱۶۷	اگر چنانچه#۱۶۳	از این جهت که#۱۵۷
بهمحض اینکه#۱۷۲	به طوریکه#۱۷۰	با وجود این#۱۶۷	اگر چنانچه#۱۶۳	از این جهت که#۱۵۷
بهمحض این که#۱۷۲	ببطوریکه#۱۷۰	با وجود این#۱۶۷	اگر چنانچه#۱۶۳	از این روی که#۱۵۸
به هر حال#۱۷۳	بطوری که#۱۷۰	با وجود این#۱۶۷	اگر چنانکه#۱۶۴	از این روی که#۱۵۸
بهر حال#۱۷۳	ببطوریکه#۱۷۰	با وجود این#۱۶۷	اگر چنانکه#۱۶۴	از این روی که#۱۵۸
ببهر حال#۱۷۳	به طوریکه#۱۷۰	بس که#۱۶۸	اگر چنانکه#۱۶۴	از این روی که#۱۵۸
به هر حال#۱۷۳	بطوریکه#۱۷۰	بسکه#۱۶۸	اگر چنانکه#۱۶۴	از این روی که#۱۵۸
به هر حال#۱۷۳	بطوریکه#۱۷۰	ببشرط آنکه#۱۶۹	اگر چنان که#۱۶۴	از این روی که#۱۵۸
ببهر حال#۱۷۳	بهمجرد#۱۷۱	بشرط آن که#۱۶۹	اگر چنانکه#۱۶۴	از این روی که#۱۵۸
ببهر حال#۱۷۳	به مجرد#۱۷۱	بشرط آنکه#۱۶۹	اگر چنان که#۱۶۴	از این روی که#۱۵۸
ببهر حال#۱۷۳	بهمجرد#۱۷۱	بشرط آنکه#۱۶۹	اگر چنان که#۱۶۴	از بس که#۱۶۰
بیآنکه#۱۷۴	به مجرد#۱۷۱	به شرط آن که#۱۶۹	باین وجود#۱۶۵	از بس که#۱۶۰
بیآن که#۱۷۴	بمحض اینکه#۱۷۲	به شرط آنکه#۱۶۹	با این وجود#۱۶۵	از بس که#۱۶۰
بی آنکه#۱۷۴	بمحض اینکه#۱۷۲	به شرط آنکه#۱۶۹	باین وجود#۱۶۵	از بس که#۱۶۱
بی آن که#۱۷۴	به محض این که#۱۷۲	ببشرط آن که#۱۶۹	با این وجود#۱۶۵	از بس که#۱۶۱
بی آن که#۱۷۴	بهمحض این که#۱۷۲	به شرط آنکه#۱۶۹	با این وجود#۱۶۵	از بس که#۱۶۱
بیآنکه#۱۷۴	به محض اینکه#۱۷۲	به شرط آنکه#۱۶۹	با اینکه#۱۶۶	از بس که#۱۶۱
تا جائیکه#۱۷۵	به محض این که#۱۷۲	ببشرط آنکه#۱۶۹	با اینکه#۱۶۶	از بس که#۱۶۱

پيوست ۴:

زنجيره املايي حروف ربط مرکب جهت رفع مشکل تنوع املايي (ادامه)

ادامه حروف ربط مرکب		
هر گاه که #۱۸۷	در نتیجه #۱۸۰	تاجایی که #۱۷۵
هر گاه که #۱۸۷	در نتیجه #۱۸۰	تاجایی که #۱۷۵
هر گاه که #۱۸۷	زیرا که #۱۸۱	تاجایی که #۱۷۵
هر گاه که #۱۸۷	زیرا که #۱۸۱	تاجایی که #۱۷۵
هر گاه که #۱۸۷	زیرا که #۱۸۱	تاجایی که #۱۷۵
همانطور که #۱۸۸	و اساساً #۱۸۲	تاجایی که #۱۷۵
همانطور که #۱۸۸	و اساسی #۱۸۲	تاجایی که #۱۷۵
همان طور که #۱۸۸	وقتی که #۱۸۳	تاجایی که #۱۷۵
همان طور که #۱۸۸	وقتی که #۱۸۳	تاجایی که #۱۷۵
همان طور که #۱۸۸	وقتی که #۱۸۳	تاجایی که #۱۷۵
همان طور که #۱۸۸	هر وقت که #۱۸۴	چنانچه #۱۷۶
همانطور که #۱۸۸	هر وقت که #۱۸۴	چنانچه #۱۷۶
همانطوریکه #۱۸۹	هر وقت که #۱۸۴	چونکه #۱۷۸
همانطوریکه #۱۸۹	هر وقت که #۱۸۴	چونکه #۱۷۸
همان طوری که #۱۸۹	هر وقت که #۱۸۴	چون که #۱۷۸
همان طوریکه #۱۸۹	هر وقت که #۱۸۴	در حالی که #۱۷۹
همان طوریکه #۱۸۹	هر وقت که #۱۸۴	در حالی که #۱۷۹
همانطوری که #۱۸۹	هر چند که #۱۸۶	در حالیکه #۱۷۹
همانطوری که #۱۸۹	هر چند که #۱۸۶	در حالیکه #۱۷۹
همانطوریکه #۱۸۹	هر چند که #۱۸۶	در حالی که #۱۷۹
همین که #۱۹۰	هر چند که #۱۸۶	در حالیکه #۱۷۹
	هر چند که #۱۸۶	در حالیکه #۱۷۹
	هر چند که #۱۸۶	در حالیکه #۱۷۹
	هر چند که #۱۸۶	در نتیجه #۱۸۰

