



ریشه یاب ماضی و مضارع از
مصدر افعال ناگذر در زبان فارسی

مجری طرح:

شاپوررضا برنجیان

عضو هیات علمی و مدیر گروه پژوهشی زبانشناسی رایانه‌ای

مرکز منطقه‌ای اطلاع رسانی علوم و فناوری

۱۳۹۰

عنوان طرح: ریشه‌یاب ماضی و مضارع از مصدر افعال ناگذر در زبان فارسی

مجری طرح: شاپوررضا برنجیان

کتاب‌آرایی، طرح‌جلد: کریم فلاح

نوبت چاپ: اول

تاریخ انتشار: ۱۳۹۰

تیراژ: ۱۰۰۰ نسخه

ناشر: مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری

با همکاری: انتشارات نوید شیراز

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.

طرح تحقیقاتی «ریشه‌یاب ماضی و مضارع از مصدر افعال ناگذر در زبان فارسی» در راستای تحقق اهداف فوق تهیه شده و مرکز منطقه‌ای اطلاع رسانی علوم و فناوری افتخار دارد که سی‌وششمین اثر خود را که حاصل تلاش فکر محققان و کارشناسان این سازمان است به زیور طبع بیاراید و تقدیم جامعه علمی و فنی بنماید.

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 مهران

رئیس مرکز منطقه‌ای اطلاع رسانی علوم و فناوری
سرپرست پایگاه استنادی علوم جهان اسلام (ISC)

فهرست مندرجات

صفحه	عنوان
۱	چکیده
۲	مقدمه
۲	روش تحقیق و پژوهش
۱۲	اشکالات اولیه برنامه ورود اطلاعات
۱۷	جامعه آماری پژوهش
۱۸	مراحل انجام پژوهش
۱۹	آزمون نظام
۲۰	پژوهش‌های انجام شده
۲۱	پسوندهای مصدر ساز
۲۴	معرفی نظام ریشه یاب ماضی و مضارع
۲۵	مزایای ساخت نظام ریشه یاب ماضی و مضارع
۲۷	نحوه عملکرد
۲۹	نتیجه‌گیری
۳۱	منابع

فهرست جداول و نمودار

صفحه	عنوان
۵	جدول ۱: لیست مصدرهای موجود در زبان فارسی
۲۳	جدول ۲: پسوندهای مصدرساز و چگونگی شکل‌گیری بن‌های مضارع.....
۲۶	نمودار ۱: قسمت‌های مختلف یک ریشه‌یاب بن مضارع از مصدر.....

سپاسگزاری

تهیه، تنظیم و تدوین این طرح بدون همکاری و یاری افراد گوناگون امکان پذیر نبود، لذا در اینجا بر خود لازم می دانم تا از تک تک آن بزرگواران تشکر و سپاسگزاری نمایم.

۱) جناب آقای دکتر جعفر مهرا، ریاست محترم مرکز منطقه ای که در تمام مراحل کار با دقت و حوصله نظارت مستمر داشته و همواره راهنما و مشوق بوده اند.

۲) اعضای محترم شورای علمی که در هنگام تصویب طرح نظرات سازنده ای ارائه نمودند.

۳) جناب آقای مهندس شمرا نی که با برنامه نویسی این طرح امکان اجرایی شدن آن را فراهم نموده اند.

۴) آقای مهدی همتی و خانم‌ها مریم ناصری، کاراندیش و معدلی که با جمع آوری اطلاعات ما را یاری نمودند.

۵) خانم هاجر صالحی که ورود اطلاعات و تایپ این گزارش را در این طرح بر عهده داشتند.

شاپور رضا برنجیان

مرکز منطقه‌ای اطلاع رسانی علوم و فناوری

گزارش نهایی طرح پژوهشی

ریشه‌یاب ماضی و مضارع از مصدر افعال ناگذر در زبان فارسی*

شاپور رضا برنجیان

عضو هیات علمی گروه پژوهشی زبان‌شناسی رایانه‌ای

چکیده:

در این طرح نمای کلی از نرم‌افزار ریشه‌یاب ماضی و مضارع از مصدر افعال ناگذر در زبان فارسی در مرکز منطقه‌ای اطلاع رسانی علوم و فناوری تهیه و ارائه گردیده و در آن عملکرد نرم‌افزار به گونه‌ای است که با ارائه مصدر فعل می‌توان به بن ماضی و بن مضارع آن مصدر دسترسی پیدا کرد. علاوه بر آن، تعدادی از افعال که بن مضارع آنها برخلاف روال معمول ساخته می‌شوند نیز بعنوان استثناء ذکر شده‌اند.

* تاریخ تصویب طرح: ۱۳۸۷/۱۰/۱

تاریخ پایان طرح: ۱۳۸۹/۹/۱

مقدمه:

با افزایش روزافزون مدارک و اسناد در شبکه جهانی وب و مراکز اطلاع رسانی الکترونیکی و کتابخانه‌های رایانه‌ای و مراکز اسناد دیجیتالی، کندی بازیابی و ذخیره سازی در عملکرد کامپیوترها و سرورها مشاهده گردید. براین اساس متخصصان علوم مختلف از جمله علوم اطلاع رسانی و کامپیوتر جهت کاهش حجم اسناد و مدارک در زمان ذخیره سازی و افزایش سرعت در هنگام بازیابی اطلاعات، راهکارهایی را ارائه و پیشنهاد کردند که یکی از این راهکارها بکار بردن ریشه‌یابها در سیستم‌های بازیابی اطلاعات بود. ریشه‌سازی اصطلاحات فرایندی است زبان‌شناختی که می‌کوشد پایه و بن هر واژه موجود در متن را مشخص کند. (مهرداد و ناصری، ۱۳۸۷)

روش تحقیق و پژوهش:

ابتدا تمامی افعال زبان فارسی از فرهنگ معین (معین، ۱۳۷۱) استخراج گردید که تعداد آنها به هشت هزار و صد مصدر می‌رسید، سپس افعال ساده از افعال مرکب جدا شده و بعد از آن افعال پیشوندی استقاقی نیز از افعال ساده مجزا گردید.

اگرچه تعدادی از افعال در این فرهنگ ضبط شده که چندان کاربردی در زبان فارسی امروزی ندارد، ولی به دلیل انجام بررسی دقیق و استخراج قوانین مورد

نیاز این طرح، تمامی آنها مورد بررسی و تجزیه و تحلیل قرار گرفت و بن‌های ماضی و مضارع آنها استخراج گردید.

سپس کلیه افعال براساس شکل مصدر و بن‌های ماضی و مضارع دسته‌بندی شدند و هر یک از افعال در گروه‌های مربوطه قرار گرفتند.

از تمامی این مصدرها تمامی افعال و مصدرهای مرکب و افعال مرده (مانند: یویدن، و رساخیدن و ...) و قدیمی حذف گردیده و تعداد ۱۴۵۲ فعل باقی ماند. بررسی‌های انجام شده، مصدرهای زبان فارسی را به شرح زیر مشخص نمود:

۱- تعریف مصدر: مصدر کلمه‌ای است که مانند فعل بر وقوع کاری یا حالتی دلالت دارد، بدون زمان و شخصی (صیغه‌های ششگانه) و مفرد و جمع. در کتاب پنج استاد آمده است مصدر از برای بیان حدوث فعلی است که به فاعلی منسوب باشد و یا شکل نامحدود فعل که در بسیاری از زبانها شکل بنیادی فعل شمرده می‌شود. (چکنی، ۱۳۸۲)

۱-۱- انواع مصدر:

۱-۱-۱- مصدر اصلی، مصدری است که از اصل به عنوان مصدر به شمار می‌رود.

۱-۱-۲- مصدر جعلی=(صناعی): مصدری است که در اصل مصدر نیست و با افزودن نشانه «یدن» و «نیدن» به آخر اسم پدید می‌آید. مانند:

ریشه‌یاب ماضی و مضارع از مصدر افعال ناگذر در ...

طلب+یدن ← طلبیدن (مدرسی، ۱۳۸۷)

۱-۱-۳- مصدر بریده = (مرخم، مخفف) ≠ (مصدر تام): مانند: گفت، شنود

(انوری و احمدی، ۱۳۸۵)

۱-۱-۴- مصدر تام: مصدری است که حذفی صورت نگرفته باشد.

۱-۱-۵- مصدر بسیط = (مصدر ساده): دارای اجزایی نیست که بتوان آنها را

جدا کرد. مانند: (آمدن)

۱-۱-۶- مصدر ذوجهین = (دوگانه): مصدرهایی که افعال آنها گاهی به

صورت لازم و گاهی به صورت متعدی بکار می‌روند. مانند: (شکستن)

(فرشیدور، ۱۳۸۲)

۱-۱-۷- مصدر لازم: مصدری است که فعل آن به مفعول بی‌واسطه نیاز

نداشته باشد. مانند (آمدن)

۱-۱-۸- مصدر متعدی: مصدری است که فعل آن علاوه بر فاعل به مفعول

بی‌واسطه نیز نیازمند باشد. مانند (زدن)

۱-۱-۹- مصدر سماعی: مصدرهایی که در اصل متعدی هستند. مانند

(خوردن)

۱-۱-۱۰- مصدر قیاسی: مصدرهایی که در اصل لازمند ولی با افزودن

«اند» یا «نیدن» به آخر بن مضارع آنها متعدی می‌شوند.

۱-۱-۱۱- مصدر متعدی دو مفعولی: در برخی از مصدرهای متعدی نیز

گزارش نهایی طرح پژوهشی

مانند مصدرهای لازم به ماده مضارع (-اندن) و (-انیدن) افزوده می‌شود. در این صورت فعل جدید علاوه بر مفعول بی‌واسطه به مفعول باواسطه نیز نیاز خواهد داشت و بدون آن دو معنی جمله ناقص خواهد بود، مانند: خورائیدن

۱-۱-۱۲- فعل کامل \neq ناقص: فعلی است که در هنگام صرف همه زمان‌ها

وجود دارد.

۱-۱-۱۳- فعل ناقص: فعلی است که در همه زمان‌ها صرف نمی‌شود.

۱-۱-۱۴- مصدر مرکب \neq ساده: مصدری است که بوسیله پیشوند از مصدر

ساده ساخته می‌شود.

با توجه به مطالب بالا، مصدرهای ساده و لازم مورد بررسی و تجزیه و تحلیل قرار گرفت. همچنین شکل‌های غیرمعمول مصدرهای مختلف (مانند: پذیرفتن مربوط به پذیرفتن) نیز حذف گردید.

در جدول شماره ۱، لیست مصدرهای موجود در زبان فارسی و چگونگی

ارتباط هر کدام آورده شده است.

جدول ۱: لیست مصدرهای موجود در زبان فارسی و چگونگی ارتباط هر کدام

مخالف	مصدر
\neq جعلی	اصلی
\neq اصلی	جعلی

ریشه‌یاب ماضی و مضارع از مصدر افعال ناگذر در ...

مخالف	مصدر
قیاسی (لازم)	سماعی (متعدی)
≠ تام	مرخم
≠ تام	بریده
≠ تام	مخفف
≠ مرخم، بریده، مخفف	تام
سماعی	قیاسی
≠ ناقص	کامل
≠ کامل	ناقص
≠ متعدی	لازم
≠ لازم	متعدی
≠ یک وجهی	ذووجهین، دوگانه
≠ مرکب	ساده
≠ ساده	مرکب

در مجموعه افعال، تعدادی با شکل‌های مختلف کاربردی، و گویش‌های متفاوت مناطق مختلف وجود دارند که آنها نیز مورد بررسی قرار گرفتند. مانند: (ستاندن، ستن، استدن).

تعدادی از افعال، گاهی به صورت کامل و گاهی به صورت ناقص و مخفف

و گاهی به صورت سابق خود، ظاهر می‌شوند، و بن‌های مضارع مشترک و یکسانی دارند که ما آنها را به دلیل اینکه هنوز در جامعه کاربرد دارند نیز به همان شکل در گروه‌های مربوطه آورده‌ایم مانند: گسستن، گسیختن، گسختن، گسلیدن=گسل

- براساس بررسی‌ها، مصدرهای مرکب و مصدرهای پیشوندی از لیست مصدرها حذف گردید. و فقط مصدرهای ساده (بسیط) باقی ماند.

- تعدادی از افعال کم بسامد که در زبان فارسی امروزی کاربردی ندارند (مانند: ورساختن و نحستن) حذف گردید.

۲-۱-۲- سپس تعیین ریشه‌های ماضی و مضارع مصدرها آغاز گردید و نتایج زیر بدست آمد.

۲-۱-۱- تمامی فعل‌های فارسی از بن‌های ماضی و مضارع ساخته شده‌اند.

۲-۱-۱- بن ماضی: بن ماضی از مصدر بدون (-ن) پایانی ساخته می‌شود (رفتن=رفت)

۲-۱-۲- بن مضارع: از فعل امر دوم شخص مفرد بدون (-ب) آغازی ساخته می‌شود مانند: (رفتن=برو=[رو])

۳- در زبان فارسی، تعدادی از واژه‌ها و تعدادی از زمان‌های افعال معمولاً از شکل‌های امری افعال ساخته می‌شوند، از این رو، از نظر زبان‌شناختی، پیدا کردن وجه امری واژه، نخستین گام در استخراج ریشه به حساب می‌آید.

ریشه یاب ماضی و مضارع از مصدر افعال ناگذر در ...

وجه امری مصدرها، با حذف دو (۲) یا سه (۳) کاراکتر از آخر مصدرها بدست می آیند. مانند: ۱-۳- و زانندن=وزان(دن)

۲-۳- ترسیدن=ترس(یدن)

۳-۳- گاهی سه حرف (-ادن) از آخر مصدر حذف می شود تا بن مضارع ساخته شود: مانند: (ایستادن=ایست ، نهادن=نه)

۴-۳- گاهی نیز ضمن حذف سه حرف آخر از مصدر، حرفی نیز به باقی مانده، افزوده می شود. مانند: حذف (-ختن) و افزودن (-ز): آویختن=آویز

۵-۳- حذف (-ودن) و افزودن (-آی): ستودن=ستای

۶-۳- حذف (-شتن) و افزودن (-ر): کاشتن=کار

۷-۳- حذف (-فتن) و افزودن (-ب): یافتن=یاب

۱-۷-۳- حذف (-فتن) و افزودن (-ف) مانند: (بافتن=باف) (شکفتن=

شکف)

۸-۳- حذف (-ستن) که به پنج شکل دیگر امکان ساختن بن مضارع وجود

دارد.

۱-۸-۳- حذف (-ستن) و افزودن (-ی): آراستن=آرای

۲-۸-۳- حذف (-ستن) بدون افزودن حرفیدن: دانستن=دان

۳-۸-۳- حذف (-ستن) و افزودن (-ه): کاستن=کاه

۴-۸-۳- حذف (-ستن) و افزودن (-ند): بستن=بند

۳-۸-۵- حذف (ستن) و افزودن (و[ی]): جستن=جوی

۴- باید توجه داشت که هر یک از این حروف متناسب با مصوت باقی مانده، پس از حذف (ستن) متفاوت می باشد. مانند دو پاراگراف زیر.

۴-۱- اگر مصوت باقی مانده (ا)[a] یا (آ)[ā] باشد، پس از آن مصوت، واج (ه)[h] افزوده خواهد شد. مانند: (خواستن=خواه)

۴-۲- اگر مصوت باقی مانده (اُ)[o] باشد، به مصوت مرکب (او)[ow] تبدیل می شود. مانند: (رستن=رو) و گاهی با مصوب بلند «و» و صامت «ی» تقابل دارد مانند: (شُست=شوی) فرشیدور، ۱۳۸۳: ۴۲۷).

بدیهی است که هر یک استثناهای خود را دارند.

چون در زبان فارسی صرف افعال توسط دو بن ماضی و مضارع (که اشکال مختلف دارند) ساخته می شوند، لذا بنظر می رسد باید ذخیره سازی به صورت مصدر انجام گیرد.

همانگونه که در زبان انگلیسی به صورت مصدر بدون (to) ذخیره می شود.

سپس گروه بندی افعال فارسی مجدداً مورد بررسی قرار گرفت و پس از آن استثناءهای هر یک از گروهها، استخراج شده و مشخص گردید.

در این میان تعدادی از افعال نیز تحت عنوان افعال ناقص استخراج گردید و همچنین به مطالبی در خصوص افعال برخورداریم که به تعدادی از آنها در زیر اشاره می شود.

ریشه‌یاب ماضی و مضارع از مصدر افعال ناگذر در ...

- بعضی از افعال ظاهراً دارای بن‌های ماضی مشترک هستند (البته بدون توجه به مصوت‌های کوتاه): مانند: رفتن=رفت ، رفتن=رفت
- بعضی نیز ظاهراً دارای بن‌های مضارع مشترک هستند (البته بدون توجه به مصوت‌های کوتاه): کندن←کن ، کردن←کن
- تعدادی نیز، خود افعال دارای شکل ظاهری یکسان می‌باشند: جستن←جستن
- تعدادی از افعال نیز دارای یک بن مضارع مشترک می‌باشند: (گسلیدن، گسستن، گسیختن، گسختن←گسل)
- پس از تهیه استنهاها، جهت تعیین خط مشی‌های پژوهشی جلسات مختلفی با متخصصین و پژوهشگران برجسته انجام گرفت و قواعد و Rule Base‌های تهیه شده پس از بحث و بررسی به تصویب رسید و به قسمت برنامه نویسی ارائه گردید تا در تهیه الگوریتم مورد استفاده قرار گیرند.
- پس از این مرحله برنامه نویس محترم جهت ورود اطلاعات اولیه، نرم افزاری تهیه نمود. (تصویر شماره ۱)

گزارش نهایی طرح پژوهشی

تعداد کل استثناء های وارد شده	تعداد استثناء 'یدن'	تعداد استثناء 'ان'	تعداد استثناء 'تن'	تعداد استثناء 'ستن'
87	12	5	5	11
	8	4	5	14
	8			
	20			

تصویر شماره ۱

از این زمان ورود اطلاعات آغاز گردید، در قسمت مصدر فعل، کلیه مصدرهای استثناء فارسی که بن مضارع آنها با هیچ یک از قواعد هم‌خوانی نداشت وارد گردید و بن مضارع آن نیز در قسمت «بن مضارع» وارد شد و در برابر هر یک، نوع آنها نیز مشخص شد. (تصویر شماره ۲) و تعداد آنها به طور خودکار در صفحه نمایش داده می‌شد.

The screenshot shows a software window with the following elements:

- Window title: ریشه‌یاب بن ماضی و مضارع از مصدر افعال فارسی
- Buttons: امکان‌ات, مصدر فعل, بن مضارع, شرح فعل
- Form fields: Input boxes for 'مصدر فعل', 'بن مضارع', and 'شرح فعل'.
- Dropdown menu: A list of verb stems including 'یدن', 'دن', 'تن', 'ذخیره', 'ادن', 'ختن', 'ودن', 'شتن'.
- Table of counts:

نوع استثنا	تعداد	تعداد استثنا "یدن"	تعداد کل استثنا های وارد شده
تعداد استثنا "ادن"	12	87	
تعداد استثنا "ختن"	4	8	
تعداد استثنا "ودن"	5	8	
تعداد استثنا "شتن"	11	20	
تعداد استثنا "ختن"	14		

Refresh button is located at the bottom left.

تصویر شماره ۲

اشکالات اولیه برنامه ورود اطلاعات:

ابتدا این برنامه به اشکالاتی برخورد که توسط متخصص کامپیوتر و برنامه‌نویس رفع گردید. در زیر به تعدادی از آنها اشاره می‌شود.

- ۱- واژه‌های تکراری مشخص نمی‌شدند و برنامه (آلارم) هشدار نمی‌داد.
- ۲- در قسمت تغییر اطلاعات هر دو کاراکتر تحت عنوان بن فعل و ریشه فعل ثبت شده بودند.

۳- (Refresh) غیرفعال بود.

۴- چنانچه بن تکراری بود بدون توجه به شکل مصدر هشدار می‌داد.

۵- واژه های ذخیره شده زیر گروه (-یدن) در آمار ثبت نمی شد.

گزارش نهایی طرح پژوهشی

پس از رفع این اشکالات توسط برنامه‌نویس و تغییرات جزئی در صفحه اصلی، تصحیحات آغاز گردید. (تصویر شماره ۳)



تصویر شماره ۳

در این صفحه با کلیک بر روی واژه «امکانات» (در قسمت راست بالای صفحه) دو گزینه، ظاهر می‌شوند، بازیابی و ویرایش اطلاعات F1 و خروج که با کلیک بر روی «خروج» از برنامه خارج می‌شویم، اما با کلیک بر روی «بازیابی و ویرایش اطلاعات» صفحه دیگری مانند «تصویر شماره ۴»

ریشه‌یاب ماضی و مضارع از مصدر افعال ناگذر در ...

بن فعل	ریشه	نوع استثنا	شوخ
آوردن	آزار	دن	...
شناختن	شناس	ختن	...
آمدن	آ	دن	...
بودن	بو	دن	...
زدن	زن	دن	...
شدن	شو	دن	...
کردن	کن	دن	...
مردن	میر	دن	...
مستند	مستان	دن	گرفتن

بن فعل: ریشه فعل: نوع استثنا:

شوخ:

تغییر اطلاعات حذف رکورد برگشت بازبای لیست

تصویر شماره ۴

ظاهر می‌شود که با کلیک بر روی هر یک از واژه‌ها می‌توان بن فعل و مضارع آن را تغییر داد و در قسمت شرح نیز، شرحی در مورد آن فعل وارد نمود. پس از آن می‌توان بر روی قسمت تغییر اطلاعات و یا حذف رکورد کلیک نمود که در این صورت صفحه‌ای مانند «تصویر شماره ۵» ظاهر شده و سوال «تغییرات ذخیره شود» نشان می‌دهد

گزارش نهایی طرح پژوهشی



تصویر شماره ۵

و در صورت کلیک بر روی «Yes»، سوال «رکورد حذف شود» ظاهر می‌شود

(تصویر ۶)

ریشه‌یاب ماضی و مضارع از مصدر افعال ناگذر در ...



تصویر شماره ۶

و با کلیک مجدد بر روی «Yes» رکورد حذف می‌شود و در صورت کلیک بر روی «حذف رکورد» سوال «رکورد حذف شود» را نشان می‌دهد (تصویر شماره ۷) و با تعیین پاسخ هر یک صفحه مجدداً به حالت اولیه خود باز می‌گردد.

گزارش نهایی طرح پژوهشی



تصویر شماره ۷

جامعه آماری پژوهش:

جامعه آماری این پژوهش را تمامی افعال فارسی موجود در فرهنگ معین تشکیل می‌دهد که توسط ۴ نفر فوق لیسانس (کارشناس ارشد) استخراج گردیدند. این افعال شامل تمامی افعال مختوم به (-دن) و (-تن) و همچنین افعال لازم و متعدی و افعال مرکب و ساده و افعال قدیمی و مرده بوده‌اند. همچنین افعال پیشوندی و اشکال تمامی مصدرهای زبان فارسی استخراج گردیدند. و همچنین شکل‌های مختلف تمامی افعال استخراج گردید: مانند: فراشتن، فراشتن، فراختن، فرازیدن، افراختن

مراحل انجام پژوهش:

طرح پژوهشی «ریشه‌یاب ماضی و مضارع از مصدر افعال ناگذر در زبان فارسی» که در مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری انجام گرفت در تاریخ ۸۷/۱۰/۱ آغاز شد و در سه مرحله انجام پذیرفت.

۱- مرحله اول:

- ۱-۱- تهیه لیست افعال فارسی از فرهنگ معین
- ۲-۱- دسته‌بندی و تعیین سه کاراکتر ماقبل آخر
- ۳-۱- تعیین ریشه‌های ماضی و مضارع افعال
- ۴-۱- تهیه قوانین و (Rule Base)ها
- ۵-۱- تجزیه و تحلیل و تهیه لیست و گروه‌بندی افعال
- ۶-۱- تعیین استثنای هر یک از گروهها
- ۷-۱- استخراج فرمولهای مورد نظر

۲- مرحله دوم:

- ۱-۲- آنالیز سیستم و برآورد نیاز کاربران
- ۲-۲- طراحی پایگاه داده‌ها
- ۳-۲- طراحی سیستم ورود اطلاعات
- ۴-۲- طراحی engine
- ۵-۲- طراحی Rule Base

۲-۶- طراحی دیکشنری

۲-۷- طراحی سیستم جستجو

۳- مرحله سوم:

۳-۱- نصب سیستم بر روی رایانه

۳-۲- ورود اطلاعات و استثناها

۳-۳- نصب Rule Base

۳-۴- آزمایش سیستم

۳-۵- تهیه گزارش نهایی.

آزمون نظام:

- آزمون نظام در یک محیط واقعی انجام گرفت و بخوبی پاسخ داد. از آنجایی که الگوریتم خاصی را طراحی کردیم، لذا این الگوریتم با سایر الگوریتم‌ها تفاوت فاحش دارد. این نظام قادر است پاسخ‌های سریعی را در کمترین زمان با حجم بالا فراهم آورد. (تصویر شماره ۸)



تصویر شماره ۸

در این تصویر که صفحه جستجو است می‌توان مصدر فعل مورد نظر را وارد نمود و با کلیک بر روی (بازیابی بن ماضی و مضارع) نتیجه جستجو در قسمت پایین صفحه نمایش داده می‌شود. علاوه بر این علاقمند بودیم تا نتایج حاصله را با یک نظام دیگر مقایسه کنیم. برای این کار با تعدادی از مدعیان تهیه ماشین ریشه ساز فارسی تماس گرفتیم، اما هیچ‌یک از آنان قادر به ارائه محصولی ملموس نشدند.

پژوهش‌های انجام شده:

بازیابی اطلاعات و ریشه‌سازی در زبان فارسی:

پژوهش در زمینه بازیابی اطلاعات^۱ و زبان فارسی چندان چشمگیر نیست. مهمترین و شناخته شده ترین مطالعاتی که در دسترس عموم می باشد، مقالاتی است که در دو کارگاه پژوهشی زبان فارسی و رایانه به تاریخهای ۱۳۸۳ و ۱۳۸۴ در دانشگاه تهران انجام شده است. (بی جن خان ۱۳۸۳)

این کارگاهها شامل مجموعه هایی از خلاصه مقالات و گاهی نیز اصل آنها در زمینه های مختلف بکارگیری توانمندی های علمی زبان فارسی در قالب اطلاعات الکترونیکی در محیط رایانه می باشد، که شرکت کنندگان تجربیات خود را با زبان فارسی و رایانه گزارش نموده اند، در مجموع دانش زبان شناختی آنها از زبان فارسی یا ضعیف یا ناکافی بوده است و این مسئله، ارزیابی و بازده ابزارهای آنها را ضعیف می سازد.

اکثر آنها از الگوریتم مشابه مانند پورتر (Porter, 1980) جهت ریشه سازی و ابزارها استفاده کرده اند و به همین دلیل دارای محسنات و معایب مشابه می باشند و در کل دارای تفاوت های جزئی از جمله تفاوت در لیست ها و تعداد قواعد و غیره می باشند. در اکثر این ابزارها بطور وضوح فقدان دانش زبان شناختی به چشم می خورد.

پسوندهای مصدرساز:

اصولاً تمامی مصدرهای فارسی یا به (-دن) و یا به (-تن) ختم می شوند.

ریشه‌یاب ماضی و مضارع از مصدر افعال ناگذر در ...

براساس بررسی‌هایی که انجام گرفت مشخص شد که جهت تعیین چگونگی ساخت بن مضارع می‌بایست براساس ۳ حرف آخر مصدر اقدام نمود که هر یک به شرح زیر تعیین می‌شوند.

- ۱- سه حرف آخر مصدرهایی که به (-یدن) ختم می‌شوند، با حذف (-یدن) بن مضارع ساخته می‌شود مانند: (خریدن ← خر، ترسیدن ← ترس)
- ۲- با حذف (-ستن) از آخر مصدرها بن مضارع ساخته می‌شود مانند: (مانستن ← مان) (لازار، ۱۳۸۴)
- ۳- با حذف (-دن) از آخر مصدرها بن مضارع ساخته می‌شود مانند: (شنودن ← شنو)
- ۴- با حذف (-ادن) از آخر مصدر، بن مضارع ساخته می‌شود مانند: (نهادن ← نه)
- ۵- با حذف (-ختن) از آخر مصدرها و افزودن (-ز) به باقی‌مانده، بن مضارع ساخته می‌شود مانند: (آویختن ← آویز)
- ۶- با حذف (-ودن) از آخر مصدرها و افزودن (-آی) به باقی‌مانده، بن مضارع ساخته می‌شود مانند: (ستودن ← ستای)
- ۷- با حذف (-شتن) از آخر مصدر و افزودن (-ر) به باقی‌مانده مصدر، بن مضارع ساخته می‌شود مانند: (کاشتن ← کار)
- ۸- با حذف (-فتن) از آخر مصدر و افزودن (-ب) به باقی‌مانده مصدر، بن

گزارش نهایی طرح پژوهشی

مضارع ساخته می‌شود مانند: (تافتن ← تاب)

این قانون دائمی نیست زیرا گاهی پس از حذف (فتن) حرف (ف) به

آخر آن افزوده می‌شود مانند (بافتن ← باف)

جدول ۲: پسوندهای مصدرساز و چگونگی شکل‌گیری بن‌های مضارع بر مبنای

مشخصه‌های قواعد صورت صوتی حروف پایانی افعال

پسوندهای مصدرساز	به (آ) ختم شود پس از حذف پسوند مصدرساز اضافه خواهد شد	به (آ) ختم شود	به (ب)، صامت ختم شود	به (ه) ختم شود	باقی مانده به صامت ختم شود
دن					
یدن					و شنیدن ← شنو
ادن	(ه) دادن ← ده				
ختن			ز		
ودن			آی		و شنودن ← شنو
شتن			ر		
فتن			ف/ب		
ستن	ی	ه	ند		[و]

بنابراین چون در ساخت بن مضارع در زبان فارسی اشکالات گوناگونی در

حذف حروف آخر و افزودن حروف جدید پیش می‌آید، لذا الزاماً استثناها

ریشه‌یاب ماضی و مضارع از مصدر افعال ناگذر در ...

افزایش می‌یابند و در صورت تهیه Rule Baseها الزاماً می‌بایست الگوریتمی تهیه کرد تا ابتدا بتواند واژه را از انتها بررسی کرده و معکوس عمل کند. در مرحله بعدی ابتدا یک حرف را شناسایی و سپس حرف دوم را و در صورت مطابقت با Rule Baseها با حرف سوم تطبیق داده شود. چون پسوندهای جمع‌ساز فارسی هیچ‌کدام بیش از ۳ حرف نمی‌باشند. لذا نیاز به بررسی بیشتر از ۳ حرف آخر کلمات نیست.

پسوندهای مصدرساز فارسی و جگونگی شکل‌گیری بن‌های مضارع بعضی از افعال در جدول شماره (۱) آورده شده است. لازم بذکر است که ساخت بن ماضی صرفاً با حذف (-ن) از آخر مصدر در تمام افعال فارسی انجام می‌گیرد.

معرفی نظام ریشه‌یاب بن ماضی و مضارع از مصدر افعال:

این نظام توانایی ساخت بن ماضی و مضارع از مصدرهای فارسی را دارد. این نظام از بخش‌های زیر ساخته شده است.

- ۱- تشخیص مصدر از بن‌های فعل فارسی
- ۲- تشخیص پسوندهای مصدرساز فارسی
- ۳- حذف پسوندهای مصدرساز و ارائه بن ماضی
- ۴- حذف پسوندها و ارائه بن مضارع

۵- تشخیص استثناهای موجود

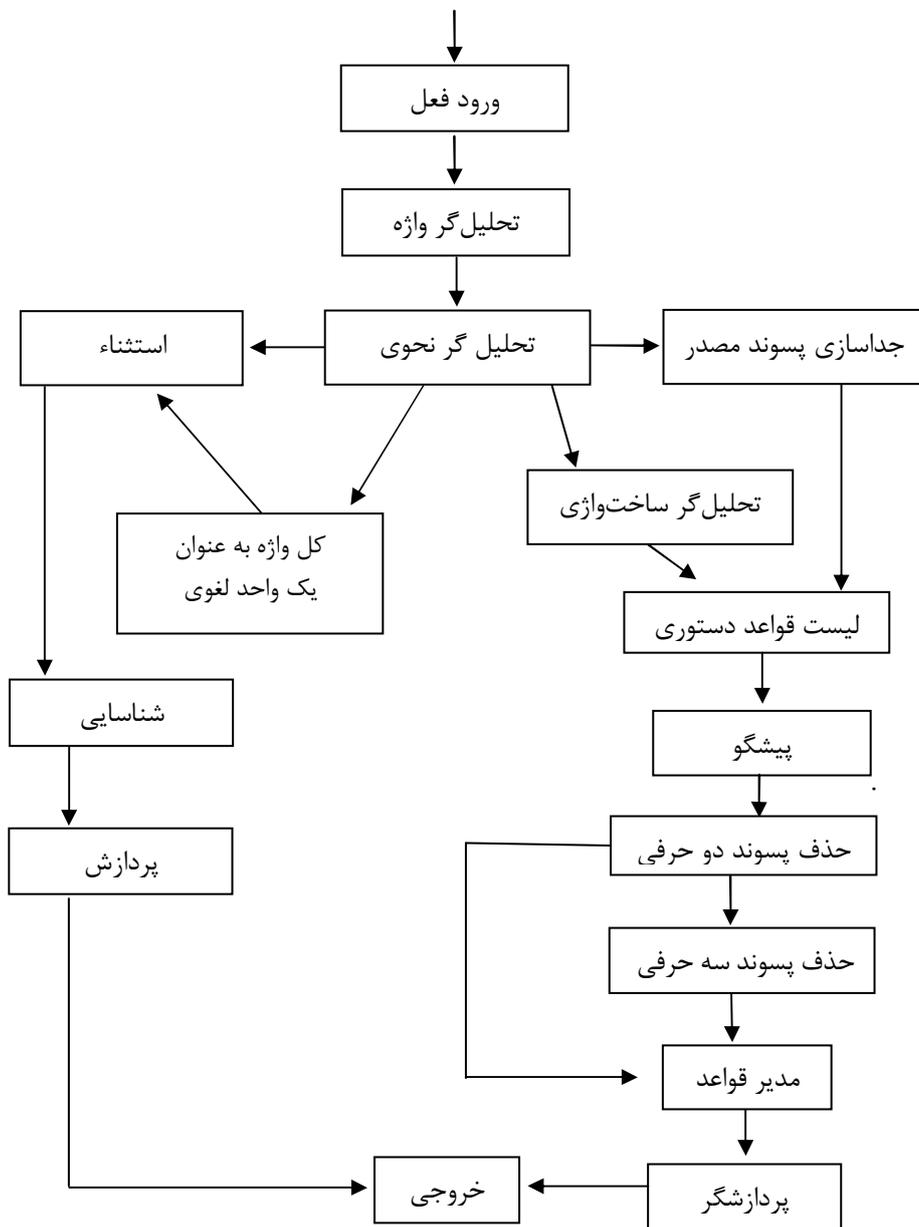
در این نظام تاکنون ۸(هشت) قاعده تعریف شده است و از مزایای این نظام آن است که کاربر می‌تواند دانش این نظام را براساس نیاز خود به روزرسانی کند و سپس به ارزیابی نتایج حاصله بپردازد. از آنجایی که از تمامی قواعد بصورت عددی پیشینه‌ای ثبت می‌شود، لذا کاربر می‌تواند از کار خود گزارش آماری نیز تهیه نماید. در این نظام ابتدا کاربر می‌تواند در هر بار یک مصدر را ارائه نموده و بن ماضی و مضارع آنرا مشاهده نماید.

مزایای ساخت نظام ریشه‌یاب ماضی و مضارع از مصدر افعال:

از این نظام استفاده‌های زیادی می‌توان برد از آن جمله می‌توان به موارد زیر اشاره نمود:

- از این نظام می‌توان برای کاهش شکل‌های مختلف یک فعل استفاده نمود.
- این نظام در دسته‌بندی خودکار متن در فایل‌های بزرگ کامپیوتری کاربرد دارد.
- از این نرم‌افزار می‌توان در مورد پژوهش در بن و ریشه کلمه‌ها جهت استفاده در نظام‌های اطلاعاتی به منظور فشرده‌سازی داده‌ها استفاده نمود.

در نمودار ۲ قسمت‌های مختلف یک ریشه‌ساز آورده شده است



نمودار ۱: قسمت‌های مختلف یک ریشه‌یاب بن مضارع از مصدر

نحوه عملکرد:

هر مصدر فعل پس از ورود، ابتدا توسط پردازشگر، مورد پردازش لغوی؛ ساختوازی و نحوی قرار می‌گیرد. پردازشگر که از روش‌های زبان پایه و مبتنی بر الگو بهره می‌برد خود دارای ۴ جزء اصلی است.

۱- تحلیل‌گر نحوی: این تحلیلگر یک تجزیه‌گر چارت عمودی است که مسئول تبدیل مصدر با لیستی از شکل درختی است که تجزیه مصدر را بر عهده دارد.

۲- تحلیل‌گر ساختوازی: در طول کار، هرگاه سیستم با یک مصدر ناشناخته برخورد کند، این قسمت از تحلیلگر سعی در تقلیل یا تشبیه‌سازی آن مصدر با یکی از اشکال از پیش تعیین شده شناخته شده می‌کند.

۳- احتساب کل واژه به عنوان یک واحد لغوی: این قسمت، چنانچه موفق به انجام دو جزء بالا نگردد کل واژه را به عنوان یک واحد لغوی در نظر می‌گیرد.

۴- جزء چهارم، استثناءهای خاص می‌باشد که در طول کار، واژه‌های خاص را که منطبق بر واژه‌های از پیش تعیین شده باشند را شناسایی و سعی در ریشه‌یابی آن می‌کند در غیر اینصورت به استثناءهای عام رفته و از شناسایی و پردازش به خروجی می‌روند.

جهت استخراج دانش لغوی فعل، ابتدا لیستی از قواعد دستور زبان که افعال بر آنها تطابق دارد تهیه می‌شود، با توجه به این فرض که تمام افعال ورودی از

ریشه‌یاب ماضی و مضارع از مصدر افعال ناگذر در ...

نظر دستوری صحیح (مجاز) هستند، لیست قواعد دستوری ارائه شده به پیشگو این امکان را می‌دهد که نقش(های) صرفی و نحوی افعال ناشناس را تعیین کند. در این مرحله فعل جدید با ویژگی‌های استخراج شده آن به قواعد دستوری ارسال می‌شود تا در نهایت مصدر توسط پردازشگر زبان طبیعی به بن مضارع و ماضی تبدیل شود.

در این پردازشگر مصدرهای فارسی در زبان فارسی به ۸ گروه زیر تقسیم می‌شوند:

۱- مصدرهایی که با حذف (-یدن) از انتهای آنها بن مضارع ساخته می‌شود:
مانند: دویدن ← دو

۲- مصدرهایی که با حذف (-دن) از انتهای آنها بن مضارع ساخته می‌شود:
مانند: آوردن ← آور

۳- مصدرهایی که با حذف پسوند (-ستن) از انتهای آنها بن مضارع ساخته می‌شود: مانند: مانستن ← مان

۴- مصدرهای که با حذف پسوند (-ادن) از آخر آنها بن مضارع ساخته می‌شود: مانند: نهادن ← نه

۵- مصدرهایی که با حذف (-ختن) از انتهای آنها و افزودن (-ز) به باقی مانده، بن مضارع ساخته می‌شود. مانند: آویختن ← آویز

۶- مصدرهایی که با حذف (-ودن) از انتهای آنها و افزودن (-آی) به

باقی مانده، بن مضارع ساخته می شود. مانند: ستودن ← ستای
۷- مصدرهایی که با حذف (-شتن) از آخر آنها و افزودن (-ر) به
باقی مانده، بن مضارع ساخته می شود. مانند: کاشتن ← کار
۸- مصدرهایی که با حذف (-فتن) از انتهای آنها و افزودن (-ب) به
باقی مانده، بن مضارع ساخته می شود. مانند: تافتن ← تاب
و در تمامی مصدرها جهت ساختن بن ماضی فقط پسوند (-ن) از انتهای
مصدر حذف می شود و بن ماضی ساخته می شود.

دسته دیگری نیز تحت عنوان استثناءها وجود دارد که نمی توان براساس
ضوابط و پیشنهادات موجه در مورد آنها قواعدی ارائه نمود. مانند: کردن ← کن و

...

در ضمن بدلیل اینکه پردازشگر تمامی واژه های مختوم به (-دن) و (-تن) را
به عنوان مصدر فعل تلقی می کند لذا، سایر واژه های مختوم به این پسوندها،
بعنوان استثناء وارد سیستم شده اند مانند: تن، تهمتن و

نتیجه گیری:

در سالهای اخیر مقالات متعددی در خصوص ریشه ساز زبان فارسی در
مجلات به چاپ رسیده است که هر یک از آنها همواره بر مبنای مباحث نظری و
براساس مقالات سایر زبان ها نوشته شده است که متاسفانه تمامی آنها بر مبنای

ریشه‌یاب ماضی و مضارع از مصدر افعال ناگذر در ...

مشاهدات علمی این نظام‌ها نبوده و مبنای علمی و عملی ندارند. اما این سیستم برای اولین بار بطور علمی و عملی در مرکز منطقه‌ای اطلاع رسانی علوم و فناوری به اجرا در آمده است. آنچه مشاهده می‌شود، بخشی از تحقیق انجام گرفته در این مرکز می‌باشد. از این نظام در شبکه درونی و سایت مرکز منطقه‌ای اطلاع رسانی علوم و فناوری استفاده می‌شود.

منابع

الف. فارسی

- انوری، حسن؛ احمدی گیوی، حسن. (۱۳۸۵). *دستور زبان فارسی*. تهران: موسسه فرهنگی فاطمی
- بی بی خان، محمود؛ قاصدی، محمد اسماعیل و پاکدل، میترا. (۱۳۸۳). *مجموعه سخنرانی‌ها و گزارش‌ها و چکیده طرح‌ها*. تهران: دانشگاه تهران.
- چکنی، ابراهیم. (۱۳۸۲). *فرهنگ دایره المعارف زبان و زبانها*. دانشگاه لرستان، نشر بهنام.
- فرشیدور، خسرو. (۱۳۸۳). *فعل و گروه فعلی و تحول آن در زبان فارسی*. تهران: انتشارات سروش.
- فرشیدور، خسرو. (۱۳۸۲). *دستور مفصل امروز*. تهران: انتشارات سخن
- قریب، عبدالعظیم؛ بهار، ملک الشعرا و سایرین. (۱۳۷۱). *دستور زبان فارسی (پنج استاد)*. به کوشش امیرالشرف الکتابی، تهران: انتشارات واژه.
- لازار، ژیلر. (۱۳۸۴). *دستور زبان فارسی معاصر*. ترجمه مهستی بحرینی، تهران: انتشارات هرمس.
- مدرس، فاطمه. (۱۳۸۷). *ازواج تا جمله*. تهران: نشر چاپار.
- معین، محمد. (۱۳۷۱). *فرهنگ فارسی (۶ جلدی)*. تهران: انتشارات امیرکبیر

ریشه‌یاب ماضی و مضارع از مصدر افعال ناگذر در ...

مهراد، جعفر؛ ناصری، مریم. (۱۳۸۷). پردازش زبان طبیعی و بازیابی اطلاعات.

شیراز، مرکز منطقه ای اطلاع رسانی علوم و فناوری، تهران: نشر چاپار.

ب. انگلیسی

Porter, M. F. (1980). *An algorithm for suffix stripping*. Program, vol. 14, No. 3, 130-137.