

چکیده

به دست آوردن سیر اشتقاق تصریف از ریشه تا کلمه، پیدا کردن کلمات مرتبط و در نهایت، تشکیل خانواده کلمات مربوط به هم، شرایط مساعدی را برای محققان پردازش زبان طبیعی فراهم می‌کند تا بتوانند از مرحله لفظ به دامنه وسیع لغات هم‌معنا، آنتولوژی، شبکه معنایی، تحلیل محتوا و در نهایت، درک معنا قدم بگذارند. در این بین، ساخت درخت سلسله‌مراتبی کلمات مرتبط با هم از ریشه (صرفی) تا برگ (کلمات متن)، اهمیت بسزایی دارد. در این مقاله، به چندین فایده از فواید «تحلیلگر صرفی نور» اشاره شده است. بعضی از این موارد، عملیاتی شده و هم‌اکنون در حال استفاده هستند و بعضی دیگر نیز نزدیک به اجرا می‌باشند. از جمله این موارد، بهره‌برداری در شناسایی خودکار سیر اشتقاق تصریف و تشکیل درخت سلسله‌مراتبی می‌باشد. عناصر داخل این درخت، همگی توسط تحلیلگر صرفی نور تولید می‌گردند. بعد از آن، عملیات دسته‌بندی و در نهایت ایجاد درخت مذکور صورت می‌پذیرد. تمام مسیرهای موجود از برگ‌ها تا ریشه، موضوعیت دارند که در این نوشتار به چند فایده از آن اشاره شده است.

کلیدواژه‌گان: تحلیلگر صرفی، اشتقاق تصریف، درخت ارتباط معنایی، پیراسته‌سازی.



شناسایی خودکار سیر اشتقاق کلمات

در زبان عربی

سید محمد دانش * 

mdanesh@noornet.net

* کارشناس متن کاوی معاونت فنی نور.

مقدمه

تحلیل صرفی و نحوی، از جمله تحلیل‌های مهم و پایه‌ای هستند که در بسیاری از پردازش‌های متنی زبان عربی مورد استفاده قرار می‌گیرند. تحلیل صرفی نیز خود در تحلیل نحوی به طور گسترده مورد استفاده قرار می‌گیرد و به همین دلیل، در زبان عربی تحلیل صرفی، از اهمیت بالایی برخوردار می‌باشد.

در حال حاضر، مرکز تحقیقات کامپیوتری علوم اسلامی از تحلیلگر صرفی نور برای نیازهای خود استفاده می‌کند. در ابتدای امر دستیابی به تمام حالات صرفی صحیح یک عبارت (کلمه) به عنوان هدف اولیه مطرح گردید؛ ولی به دلیل استفاده گسترده و فواید جانبی دیگری که بر این سامانه متصور است، بر آن شدیم تا در این مقاله به چند نمونه از این فواید اشاره نماییم که می‌توانند در پروژه‌های مرکز تحقیقات کامپیوتری علوم اسلامی و سرعت بخشی به بعضی از آنها مورد بهره‌برداری قرار گیرند. ارائه مجموعه کاملی از خصوصیات صرفی توسط برنامه، قابلیت بهره‌برداری‌های بیشتر را برای کاربردهای متفاوت فراهم می‌نماید. جداسازی پیشوندها و پسوندها، کمک به شناسایی اغلاط احتمالی متن، استفاده در تحلیل نحوی، شناسایی و پیشنهاد ریشه‌های احتمالی و نیز بهره‌برداری در سیر اشتقاق تصریف، از جمله این فواید می‌باشد که در ادامه به آنها اشاره می‌کنیم.

۱. جداسازی (۱) پیشوندها و پسوندها

پیراسته‌سازی (۲)، یکی از عملیات‌های رایج و پرفایده در زمینه پردازش هوشمند متون می‌باشد که از اهمیت بالایی برخوردار است. یکی از خروجی‌های جنبی تحلیلگر صرفی نور، این است که قبل از اینکه وارد مرحله تحلیل صرفی عمیق کلمه شود، تمام حالات جداسازی پیشوندی و پسوندی را برای کلمه ورودی انجام داده، فهرست تمام حالات احتمالی را ارائه نماید. از آنجا که بعضی از پیشوندها و پسوندها میان اسم، فعل و حرف مشترک بوده و یا میان دو دسته از آنها مشترک هستند و از طرف دیگر، بعضی از آنها مخصوص یک دسته می‌باشد، در مرحله اولیه و بدون مشخص نمودن کاربر حالات ارائه‌شده توسط برنامه متنوع بوده و ممکن است کاربر به راحتی نتواند به حالت مورد نظر خود دست پیدا کند؛ اما اگر کاربر نوع کلمه را وارد نماید، این حالات به مراتب کاهش پیدا خواهد نمود.

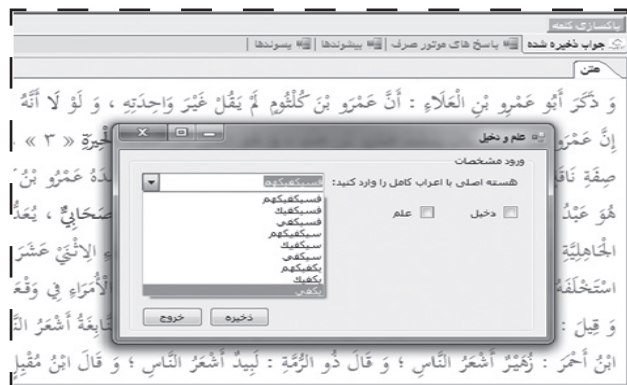
بنا به وارد نمودن نوع تفکیک اسمی یا فعلی و همچنین عمق جداسازی، می‌توان به تفکیک دلخواه دست پیدا نمود. از این قابلیت علاوه بر جداسازی پیشوندها و پسوندها و دستیابی به میانوند مورد نظر، می‌توان در مواردی که کاربر قسمتی از کلمه‌ای را وارد می‌کند، بقیه حالات را به عنوان پیشنهاد و قبل از وارد نمودن به او ارائه نمود که در صورت موجود بودن

میانوند مورد نظر، آن را انتخاب نموده، بدین صورت، عملیات وارد کردن کلمه سرعت بهتری پیدا کند و در صورت نادرست بودن قسمت وارد شده فعلی، آن را برایش اصلاح کرده، درست آن را جایگزین نمود. در شکل ذیل، تفکیک کلمه «فسیکفیکهم» نشان داده شده است که به کاربر در انتخاب کلمه مورد نظر کمک می‌کند.

با کمک قاعده مشهور «سألتمونیها» می‌توان پا را فراتر نهاد و حتی حروف «أتین» را از افعال و اسما جدا نمود تا دسته‌بندی‌های جامع‌تری داشته باشیم؛ برای مثال، کلمات «یتقابل» و «مقابل» و «مقابلک» را در نظر بگیرید که دوتای اول فعل و دوتای بعدی اسم می‌باشند. در مرحله اول، با جداسازی پیشوندها و پسوندها دو دسته با محوریت «یتقابل» و «مقابل» تشکیل می‌گردد که هر کدام هم دو کلمه از این چهار کلمه را پوشش می‌دهند؛ ولی اگر روند تفکیک را در سطح قاعده «سألتمونیها» ادامه دهیم، یک دسته با محوریت «قابل» تشکیل می‌گردد که هر چهار کلمه را درون خود جای داده است و به عبارت دیگر، با تجرید بیشتر می‌توانیم به یک میانوندی برسیم که جامعیت بیشتری داشته باشد. گفتنی است، از این قابلیت در موتور میانوند به شکل گسترده در ابزار ریشه مشتق استفاده شده است که با یک میانوند، حداکثر کلمات پوشش داده می‌شود و کاربر با انتخاب یک ریشه، ریشه مورد نظر را برای همه آنها انتخاب می‌نماید که در افزایش سرعت زدن ریشه تأثیر بسزایی دارد.

۲. شناسایی اغلاط احتمالی (۳) متن

از آنجا که روند تحلیل موتور صرف با فرض صحیح بودن کلمات موجود در متن ورودی صورت می‌پذیرد، به همین دلیل، اگر تحلیلگر صرفی نتواند برای کلمه‌ای



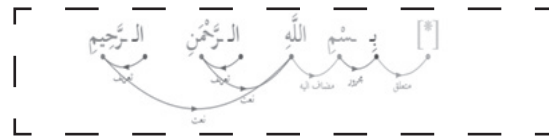
جوابی ارائه نماید، فارغ از اینکه عدم وجود جواب ممکن است به جهت ضعف و نقص تحلیلگر صرفی باشد، خود می‌تواند با یک احتمال نسبتاً بالایی وجود اشتباه برای آن کلمه را بیان نماید. این اشتباه ممکن است از چند دسته مختلف: اشتباهات املائی(تایپی)، اشتباهات ساختاری(صرفی)، به هم چسبیدگی کلمات و از هم گسستگی اجزای یک کلمه باشد. کلمات با اشتباهات ساختاری مثل «نبطال» یا «توصیف»، از نظر زبان‌شناسی نیز دارای اهمیت بالایی است. از این قابلیت می‌توان به مقدار قابل

توجه در تصحیح اغلاط متون استفاده نمود و روند تصحیح را سرعت بخشید.

۳. استفاده در تحلیل نحوی

از آنجا که شناخت جایگاه کلمه از نظر نحوی، بر مشخص بودن شرایط صرفی کلمه مبتنی است، به همین دلیل، لازم است قبل از شروع نتیجه‌گیری‌های نحوی، از لحاظ صرفی عملیات شناسایی انجام شده باشد و به عبارت بهتر، تحلیل نحوی سرعت بهتری پیدا می‌کند؛ برای مثال، فاعل یا مفعول جمله حتماً باید اسم باشند

و یا ترکیب اضافی جار و مجرور همیشه از یک حرف و اسم بعد از آن تشکیل می‌گردد. یکی دیگر از موارد تکرار در متن، ترکیبات مضاف و مضاف الیه می‌باشد که باید از دو اسم تشکیل گردد و یا شناسایی جملات فعلیه متوقف بر این است که جمله با فعل شروع گردد که همه اینها، از تحلیل صرفی استنتاج می‌گردد. در حال حاضر، از خروجی تحلیل صرفی در تحلیل نحوی قرآن کریم استفاده شده است که در نوع خود بی‌نظیر می‌باشد و در پایگاه متن نور قابل دسترسی می‌باشد.



| | | |
|---|-----|------|
| ب | حرف | باز |
| س | اسم | هسته |
| م | اسم | هسته |
| ل | اسم | هسته |
| ه | حرف | باز |
| ر | اسم | هسته |
| ا | حرف | باز |
| م | اسم | هسته |
| ی | حرف | باز |
| م | اسم | هسته |
| ه | حرف | باز |
| ر | اسم | هسته |
| ا | حرف | باز |
| م | اسم | هسته |

۴. شناسایی و پیشنهاد ریشه‌های احتمالی یکی دیگر از قابلیت‌های تحلیلگر صرفی نور، استفاده در عملیات تعیین ریشه برای کلمات متن می‌باشد. روند کار این‌گونه است که می‌توان از ویژگی ریشه خروجی برنامه صرف به طور بسزایی در شناسایی ریشه کلمات بهره برد. این نیاز در خصوص کلمات معتل، مقلوب و یا مهموز، بیشتر خودنمایی می‌کند؛ چراکه در این نوع کلمات به دلیل اعمال قواعد اعلال، مهموز، تخفیف، ادغام و ابدال، ساختار کلمه دستخوش تغییر و دگرگونی می‌شود. این امر باعث می‌گردد که تشخیص ریشه برای محقق مشکل گردد؛ برای مثال، کلمه «عاد» می‌تواند هر سه ریشه «عدد»، «عدو» و «عود» را داشته باشد و یا هر کدام از ریشه‌های «بور»، «برو»، «بری»، «ب» و «بر» می‌تواند ریشه کلمه «بار» باشند که تشخیص بعضی از آنها برای کاربر

یا مشکل بوده و بعضی دیگر هم حقیقتاً غیر قابل تشخیص هستند که در اینجا با کمک برنامه می‌توان تمام ریشه‌های احتمالی را پیدا کرده، به کاربر ارائه نمود. با داشتن تعداد تکرار ریشه‌های موجود، می‌توان ریشه‌های مستعمل و غیرمستعمل را برای هر کلمه در صورت تعدد ریشه مشخص کرد و از طرف دیگر، در صورت نقص ریشه‌های استعمالی، آن را ترمیم نمود.

۵. کمک به موتور اعراب (۴)

از آنجا که موتور صرف قابلیت کار با متون با اعراب و بدون اعراب دارد، می‌توان از خروجی آن جهت ائقان و در عین حال، برطرف نمودن اشکالات احتمالی جواب پیشنهادی موتور اعراب‌گذار استفاده کرد.

۶. استفاده از بانک‌های جانبی برنامه

این بانک‌ها با زحمت زیاد و تلاش همکاران گروه ادبیات و متن‌کاوی، از میان کتب

مختلف استخراج و جمع‌آوری شده است. بانک‌های اطلاعاتی مذکور عبارت‌اند از: بانک ادات اسمی و حرفی، بانک اسامی جامد، جمع‌های مکسر، افعال غیرمتصرف، ادات متصل، صفات مشبیه، اعلام تک‌کلمه‌ای و بانک ریشه‌های مستعمل زبان عربی، مصادر ثلاثی و رباعی مجرد. این منابع اطلاعاتی، در نوع خود بسیار ارزشمند است که به تدریج جمع‌آوری شده و غنی گردیده‌اند.

۷. آماده‌سازی پیکره (۵)های نشان‌گذاری

شده

یکی دیگر از استفاده‌هایی که از تحلیلگر صرفی می‌شود، کمک به زبان‌شناسان خبره ادبیات عرب برای تهیه پیکره‌های رفع ابهام‌شده و آماده‌سازی دیتای آموزش برای برنامه رفع ابهام صرف می‌باشد. برنامه رفع ابهام صرف، جواب‌های با استعمال بیشتر را به ترتیب اولویت‌بندی می‌کند. در حال حاضر، این برنامه با زحمت

| تعداد رکورد | نام بانک | نوع بانک اطلاعاتی |
|-------------|------------------|-------------------|
| ۶۳ | ادات متصل | اسمی |
| ۲۰۰۰۰ | اعلام | |
| ۵۰ | پسوند اسمی | |
| ۳۶۸ | ادات اسمی | |
| ۱۹۲۲۱ | جوامد | |
| ۵۳ | مصادر ثلاثی مجرد | |
| ۱۸۰۶۷ | جمع‌های مکسر | |
| ۱۲۸ | ادات حرفی | |
| ۱۸ | پیشوند فعلی | |
| ۴۳ | پیشوند اسمی | |
| ۳۳ | پیشوند حرفی | |
| ۶۷ | افعال غیرمتصرف | فعلی |

صرف، استفاده در به‌دست آوردن سیر اشتقاق تصریف و روند ساخته‌شدن کلمه از ریشه تا کلمه مورد نظر می‌باشد. ارائه سلسله‌کلمات مرتبط با کلمه ورودی، در اصطلاح، ریشه‌یابی تصریفی (۶) نامیده می‌شود. روند کار بدین صورت است که زنجیره این کلمات مرتبط با توجه به تعریف ارائه‌شده، از نزدیک‌ترین سطح شروع گردیده، تا دورترین سطح که همان ریشه کلمه است، ادامه پیدا می‌کند.

در سامانه تحلیلگر صرفی، تمام شقوق و حالات ریشه‌های معتبر زبان عربی به طور بالقوه وجود دارند و برای هر دسته از کلمات، اعم از فعل و اسم، توانایی تولید و ایجاد وجود دارد. در حال حاضر، فقط حالاتی که متناسب با کلمه ورودی هستند،

توسط برنامه تولید می‌گردد و حالات مورد نیاز، به صورت هوشمند تولید می‌شوند. از آنجا که مقدمات تولید زنجیره سیر اشتقاق تصریف، همگی در تحلیلگر صرفی موجود بود و فقط نیاز به تولید یک‌ایک عناصر این زنجیره داشت، به همین دلیل، در کنار سیر تحلیل صرفی برای کلمه، این زنجیره نیز تولید می‌گردد.

اگر ریشه صحیح و سالم نباشد، لازم است که برای تک‌تک عناصر، عملیات اعلال، ادغام، تخفیف و مضاعف نیز انجام شود که به طور متوسط، به ازای هر کلمه ورودی، ۱۰ کلمه مورد تحلیل صرفی قرار می‌گیرد. برنامه در این باره به گونه‌ای نوشته شده است که ابتدا همه عناصر به طور مجزا ساخته می‌شوند و در نهایت، با هر ترتیب دلخواهی که لازم باشد، در کنار هم قرار گرفته و به کاربر ارائه می‌شوند. این قابلیت وجود دارد که تعداد و یا ترتیب آنها عوض شده، کاربر می‌تواند متناسب با نیاز خود، زنجیره مورد نظر را دریافت نموده، مورد بررسی قرار دهد.

برای مثال، روند اشتقاق تصریفی فعل

متفاوت آمده باشد. در این صورت، لازم است برنامه بتواند تمام حالات صحیح را شناسایی نموده، از اشتباهات آن تفکیک نماید. یکی از این حروف، همزه می‌باشد که با پایه کرسی‌های مختلف در متن ظاهر می‌گردد؛ به طور مثال، کلمات «ءادم، آدم، أولئک، أولئک، شیئا، شیء، لئن، لئن، علماءهم، علمائهم، علماءهم»، از جمله مواردی هستند که برنامه می‌تواند تمام آنها را شناسایی نماید.

۹. استفاده در آموزش علم صرف زبان عربی
یکی دیگر از قابلیت‌های جنبی برنامه صرف، این است که می‌تواند در آموزش علم صرف به زبان‌شناسان، محققان، دانشجویان و طلاب علوم دینی کمک شایانی کند. اگر برنامه‌ای مناسب با طراحی واسط کاربر خوب طراحی گردد، کاربران می‌توانند خروجی‌های درخواستی خود را مشاهده نمایند که در خصوص قواعد اعلال، ادغام و یا ابدال مفید خواهد بود.

۱۰. سیر اشتقاق تصریف
یکی از مهم‌ترین کاربردهای برنامه تحلیل

همکاران گروه متن‌کاوی و مساعدت زبان‌شناسان خبره گروه پژوهش انجام شده است. سرعت‌بخشی در تهیه و توسعه این پیکره‌های رفع ابهام‌شده و افزایش دقت آن، از جمله بهره‌برداری‌هایی بود که تحلیلگر صرفی ارائه کرد.

نفس تهیه این پیکره‌های رفع ابهام‌شده با توجه به روند پرچالش و حجیمی که دارد، ارزشمند بوده و بعضی از مؤسسات مشابه، از آن به عنوان دستاوردهای انحصاری خود استفاده می‌نمایند. هم‌اینک، دیتای رفع ابهام صرف، دارای ۵۲۳۹۹۲ رکورد می‌باشد که ۴۸۷۷۱۶ رکورد آن، یعنی معادل ۹۳ درصد رفع ابهام شده‌اند. تعداد کل میانوندها، پیشوندها و پسوندهای رفع ابهام‌شده، ۴۸۹۶۳۰ عدد و تعداد کل ویژگی‌های رفع ابهام‌شده، حدود ۴۶۵۱۴۸۵ مورد می‌باشد که در نوع خود بی‌نظیر است.

۸. مدیریت رسم‌الخط‌های مختلف
برای بعضی حروف در زبان عربی، رسم‌الخط‌های مختلفی وجود دارد که همه آنها صحیح بوده و ممکن است در یک متن یک کلمه با چند رسم‌الخط

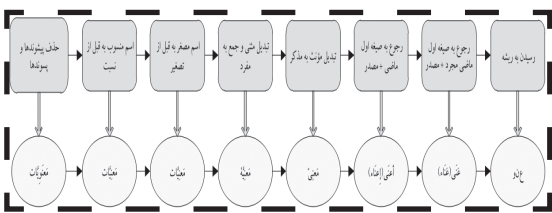
```

<Phrase Entry="سنتصّر">
<Ans number="1">
<word Seg="1" Slices="ن" Entry="ن" Affix="بیرون" Pos="حرف" Case="فح بر مبنی" Kol="حرف" />
</Ans>
<word Seg="2" Slices="تصّر" Entry="تصّر" Affix="سهه" Pos="فعل" Lemma="التصّر" />
<word Seg="3" Slices="ن" Entry="ن" Affix="تصیر" Pos="حرف" Case="فح بر مبنی" Kol="حرف" />
</Ans>
</Phrase>
    
```

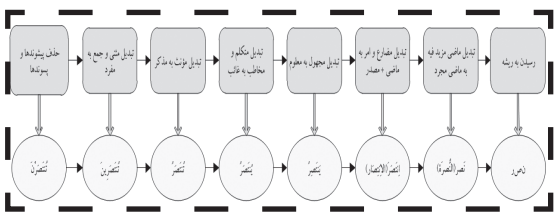
«سنتصّر» به صورت «سنتصّر»،
 تُنصّر، تُنصّرین، تُنصّر، يُنصّر، يُنصّر،
 اُنصّر، اُنصّار، نُصّر، نُصور، نُصرة، ن صر و
 برای اسم «بمعنویاته» به صورت «معنویات،
 معنیات، معنیة، معنی، أعنی، إعناء، عناء، عنو،
 عناء، عنوة، عنو، عن و» می‌باشد.

شده است، این سیر اشتقاقی ارائه می‌گردد
 که می‌توان از هر کدام از کلمات داخل این
 زنجیره در مراحل مختلف برنامه‌های
 متعدد استفاده نمود.

برای فعل «سنتصّر» در تصویر مقابل
 نشان داده شده است. همان طور که
 ملاحظه می‌کنید، برنامه تحلیلگر صرفی
 برای کلمات به صورت خطی، (۷) سیر
 اشتقاقی تصریف را ارائه می‌کند که از
 استم(میانوند) کلمه شروع شده و به ریشه
 منتهی می‌گردد. برای کل کلمات ورودی،
 بر اساس تعریفی که بسته به نوع کلمه ارائه



شکل ۲: شیوه محاسبه اشتقاق تصریفی برای اسما «بمعنویاته»



شکل ۱: شیوه محاسبه اشتقاق تصریفی برای افعال «سنتصّر»

با کنار هم قرار دادن این زنجیره‌های
 خطی، زمینه کاملاً مناسبی برای ایجاد
 ساختارهای سلسله‌مراتبی و درختی فراهم
 می‌شود و فایده اصلی کار ما در این ساختار
 به‌خوبی مورد بهره‌برداری قرار می‌گیرد.
 ریشه صرفی کلمه می‌باشد. در این مرحله،
 با کمک فنون دسته‌بندی و مرتب‌سازی

داده گروه‌ها و به عبارت دیگر، درخت‌های
 مختلفی با محوریت ریشه تشکیل می‌گردد
 که با توجه به اینکه فرزندان ریشه که
 بزرگ این دسته نامیده می‌شود، چه باشند،
 به صورت مرتبه‌به‌مرتبه و سطح‌به‌سطح
 جلو رفته و درخت فرزندان و نوه‌ها و به
 همین ترتیب، سایر نسل‌های دیگر آن مورد
 جست‌وجو قرار گرفته و هر کلمه در درخت

با کنار هم قرار دادن این زنجیره‌های
 خطی، زمینه کاملاً مناسبی برای ایجاد
 ساختارهای سلسله‌مراتبی و درختی فراهم
 می‌شود و فایده اصلی کار ما در این ساختار
 به‌خوبی مورد بهره‌برداری قرار می‌گیرد.
 ریشه صرفی کلمه می‌باشد. در این مرحله،
 با کمک فنون دسته‌بندی و مرتب‌سازی

| | | | | | | | | | | | | |
|----|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1 | فعل | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر |
| 2 | فعل | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر |
| 3 | فعل | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر |
| 4 | فعل | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر |
| 5 | فعل | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر |
| 6 | فعل | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر |
| 7 | فعل | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر |
| 8 | فعل | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر |
| 9 | فعل | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر |
| 10 | فعل | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر |
| 11 | فعل | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر | سنتصّر |

```

<Root root="نصیر">
  <Madkhal Word="تصصّر" />
  <Madkhal Word="تصصّرین" />
  <Madkhal Word="تصصیر" />
</Root>
    
```

شکل صفحه بعد، حالت باز شده‌تر این
 درخت نمایش داده شده است.

در شکل دیگر نیز هم سلسله‌مراتب پایین‌تر
 از این درخت را نیز مشاهده می‌کنید:

باعث می‌گردد کاربر به دامنه وسیع‌تری از کلمات دسترسی پیدا کند و در نتیجه، بررسی و تصمیم‌گیری‌هایش با احاطه بیشتری صورت خواهد گرفت؛ برای مثال، با جست‌وجوی کلمه «بمساجد»، علاوه بر «مساجد»، «بمساجده»، «بمساجدکم» و «مساجدهم»، کلمات «مسجد»، «سجده»، «سجود» و «سجّاد» و سایر ترکیبات پیشوندی و پسوندی آنها نیز در دامنه بررسی قرار می‌گیرند.

۳. جست‌وجوی مدیریت‌شده: اگر تمام کلمات داخل متن مورد استفاده قبل از انجام جست‌وجو توسط تحلیلگر صرفی تگ خورده و رفع ابهام هم شده باشند، در مواردی که کلمه مورد جست‌وجو بین اسم، فعل و حرف و یا حداقل دوتای از آنها مشترک باشند، کاربر را قادر می‌سازند که کلمه را در دامنه‌های مختلف اسمی، فعلی و یا حرفی مورد جست‌وجو قرار دهد؛ به عنوان مثال، کلمه «علی» بین هر سه دسته مشترک می‌باشد؛ (عَلِيّ) فعل، (عَلِيّ) اسم و عَلِيّ (حرف)). در حال حاضر، کاربر با جست‌وجوی این کلمه به سوی هر سه دسته از این کلمات در متن سوق داده می‌شود که با فرض فعال بودن جست‌وجوی پیشرفته، تعداد کلمات پیشنهادی به مراتب زیاد بوده و کاربر را دچار سردرگمی می‌نماید؛ اما با این قابلیت کاربر می‌تواند فقط «علی»های فعلی را مورد بررسی قرار دهد؛ برای مثال، کلمه «فی» میان فعل و حرف مشترک است که تعداد رخدادهای حرفی آن در متن به مراتب بیشتر از رخدادهای فعلی است و اگر کاربر بخواهد به حالات فعلی دسترسی پیدا کند، فرآیند جست‌وجو زمان‌بر، کم‌فایده و پرهزینه خواهد بود؛ ولی به کمک این قابلیت می‌تواند به موارد فعلی به طور مستقیم دسترسی داشته باشد. به دلیل ارائه شدن ویژگی‌های متعدد صرفی برای هر کلمه، می‌توان عملیات جست‌وجو را بر

اساس هر کدام از این ویژگی‌ها متمرکز نمود که در مثال فوق، جست‌وجو بر اساس نوع کلمه که در ویژگی Pos خروجی تحلیل صرف بود، صورت می‌گرفت؛ به عنوان مثال دیگر، می‌توان جست‌وجو را بر ریشه کلمات، یعنی Root خروجی تحلیل صرفی متمرکز نمود؛ برای مثال، کلمه «عاد» در حالت فعلی از ریشه «عود»، و در حالت اسمی از ریشه «عدد» قابل تصور است. حال اگر کاربر در هنگام جست‌وجو ریشه مورد نظر خود را هم وارد نماید، می‌تواند به طور مستقیم به کلمه «عاد» مورد نظر خود دسترسی پیدا نماید. در موارد لازم می‌توان حتی با تلفیق این ویژگی‌ها، فرآیند جست‌وجو را با دقت بسیار بالایی مدیریت و نوبری کرد. در مثال جست‌وجوی «علی»، کاربر می‌تواند با ترکیب نوع کلمه و ریشه به طور مستقیم، به «علی»هایی که فعل بوده و از ریشه «علو» هستند، دست پیدا کند و یا اینکه به «علی»هایی که از ریشه «علی» هستند، رجوع نماید.

۴. کمک به محققان جهت تدوین لغت‌نامه، از جمله کاربردهای دیگر سیر اشتقاق تصریف می‌باشد. روش کار این است که محقق به ترتیب از خود کلمه تا ریشه جلو رفته و در این پیشروی، هر کلمه‌ای که موضوعیت قرار گرفتن به عنوان یک مدخل در لغت‌نامه را داشته باشد، انتخاب گردیده، در جایگاه خود قرار داده می‌شوند و باعث می‌شود تا این تدوین، با نظم و دقت بهتر و سرعت بالاتری صورت پذیرد.

۵. از جمله کاربردهای مفید این ویژگی، استفاده در انتقال کاربر به نزدیک‌ترین کلمه موجود در لغت‌نامه (قاموس) می‌باشد. در این حالت، کاربر در بررسی کلمه مورد نظر خود به یکباره به سطوح دورتر و کمتر مرتبط با کلمه ورودی منتقل نشده و موجبات سردرگمی او ایجاد نمی‌گردد و تنها در صورت نیاز و درخواست خود کاربر است

که مرحله‌به‌مرحله به سطوح کمتر مرتبط که آخرین آنها ریشه کلمه است، منتقل می‌شود و دامنه بررسی کاربر به اختیار خود او گسترش پیدا می‌نماید.

کاربردهای سیستم ریشه‌یابی تصریفی، به این موارد خلاصه نمی‌شود. به طور کلی، با استفاده از این سیستم می‌توان به «غنی‌سازی انواع داده‌های متنی» پرداخت و با بهره‌گیری از «درخت‌های ارتباط معنایی»، از این متون غنی‌شده در کاربردهای سطح بالایی، از قبیل سیستم‌های: «تحلیل محتوا»، «مشابه‌یاب متون» و «معاجم لفظی و موضوعی» بهره برد. بررسی و توضیح موارد اشاره‌شده، به تفصیل و توضیح بیشتر نیاز دارد که در این مقام نمی‌گنجد و در کارهای تحقیقاتی پیش رو، با جزئیات بیشتر به آنها پرداخته خواهد شد.

پی‌نوشت‌ها:

1. Tokenization.
2. Stemming.
3. Error Detection.
4. Diacritizator.
5. Corpus.
6. Lemmatization.
7. Linier. ■