جستجو در اينترنت با استفاده از زبان طبيعی فارسی

محسن كاهاني
kahani@um.ac.ir
استاديارگروه مهندسي كامپيوتر- دانشگاه فردوسي مشهد

چكيده:

در اين مقاله استفاده از زبان فارسي محاوره اي در جستجوهاي اينترنتي مورد بررسي قرار مي گيرد و نتايج حاصل ازانجام يك پروژه كاربردي تشريح خواهد شد. هدف از اجراي اين پروژه دريافت يك جمله بزبان فارسي از كاربر است كه شامل توضيحاتي دربارة موضوع يا موضوعاتي كه وي مايل به يافتن اطلاعاتي در آن خصوص مي باشد. سيستم اين جمله فارسي را پردازش مي نمايد ، كلمات اضافي را حذف و كلمات كليدي را بصورت تركيبي ازعبارات جبري از آن استخراج مي كند. عبارت استخراج شده سپس به يك موتور جستجو ارسال مي نمايد و نتيجه حاصل به كاربر ارائه مي گردد.

كلمات كليدي: جستجوي اينترنتي؛ پردازش زبان طبيعي فارسي، هوش مصنوعي

۱- مقدمه

گسترش روز افزون اينترنت و شبكه جهاني وب و تغيير آن از يك محيط آكاديمك به يك محيط بيشتز تجاري باعث افزايش تعداد كاربران اينترنت و تنوع آنها شده است.گرچه حجم عمده اي صفحات وب بزبانهاي غير فارسي مي باشد،تعداد سايت هاي اطلاع رساني، سرگرمي و وبلاگهاي بزبان فارسي نيز بسرعت در حال افزايش است. يافتن اطلاعات مورد نياز از درون اين محيط بزرگ بعضا به يافتن يك سوزن در ميان يك كوه از كاه مانند مي گردد.

موتورهاي جستجو با امكانات متنوع و نسبتا زيادي كه در اختيار كاربران مي گذارند پيدا كردن مطالب مورد نياز را ساده تر مي نمايند. با اينحال استفاده بهينه از اين امكانات نيازمند آشنايي كاربران با اصول منطق و جبر گزاره هاست كه طيف عظيمي از كاربران وب با آنها آشنايي ندارند. جهت رفع اين مشكل برخي از موتورهاي جستجو امكان دريافت جملات بصورت زبان محاوره اي انگليسي را فراهم آورده اند تا كاربران عادي بتوانند براحتي مطالب مورد نياز خود را جستجو نمايند.

با توجه به افزايش كاربران غير حرفه اي اينترنت در ايران و وارد شدن شركتهاي تجاري در اينترنت، نياز به ايجاد يك سيستم دريافت جملات فارسي بصورت محاوره اي، حذف كلمات زايد، استخراج كلمات اصلي و تبديل آن به يك عبارت منطقي قابل ارائه به يك موتور جستجو احساس ميگردد. در

اين مقاله ضمن پرداختن به اين موضوع، ساختار سيستمي كه جهت اينكار طراحي شده است تشريح مي شوند و برخي نتايج حاصل از اجراي آن بيان مي گردد.

در اين مقاله ابتدا به مفاهيم كلي پردازش و فهم زبانهاي طبيعي اشاره مي رود. سپس به مسائل خاص زبان فارسي در اين خصوص پرداخته شده و ساختار سيستم پيشنهادي تشريح ميگردد. در نهايت پياده سازي سيستم،نتايج حاصله و كارهاي آتي جهت بهبود سيستم بيان خواهد شد.

۲- مفاهيم پردازش زبان طبيعي

«پردازش زبان طبيعي» (NLP:Natural Language Processing) يكي از مقوله هاي مطرح در شاخه وسيع و پركاربرد «هوش مصنوعي» در زمينه هاي علوم كامپيوتر است ، كاربردهاي هوش مصنوعي امروزه آنقدر وسيع شده است كه گويي يكي از اصلي ترين شاخه هاي مطرح علم كامپيوتر است روند استفاده از كامپيوتر ها و نرم افزارها اكنون به سويي پيش مي رود كه هرچه راحت تر قابل استفاده توسط كاربر يا به اصطلاح "User Friendly" باشد . اين هدف هنگامي محقق مي شود كه بتوان با ماشين براحتي ارتباط برقرار كرد . از آنجايي كه زبان ابزار مهم ارتباطي بين انسان ها است چه بصورت استفاده هاي نوشتاري و چه گفتاري، ايجاد قابليت درك و توليد زبان طبيعي در ماشين ما را به سوي ارتباط مفيد و راحت با آن سوق مي دهد .

اهميت مقوله «پردازش زبان طبيعي» نيز از همين جا سرجشمه مي گيرد. NLP موضوعي است از تحقيق پيرامون « زبان طبيعي»و رايج بين انسان ها (مفاهيم زبان شناسي) آغاز مي شود و به چگونگي ايجاد درك و فهم زبان در ماشين ختم مي شود. NLP به دو شاخة عمده تقسيم مي گردد : پردازش گفتار و پردازش نوشتار. در پردازش گفتار علاوه بر درك زبان، مقولة درك صحبت نيز وجود دارد و كار مفصل تر و پيچيده تر مي شود.

در حال كلي برخي از مفاهيم و علومي كه در اين راستا مطرح هستند عبارتند از[1]:

- Phonological : تشيخص اصوات در لغات
- Morphological : علم لغت و « واژك ها » (Morphemes) كه كلمات را تجزيه كند .
- Syntactic : شناخت نحوي و گرامري
- Semantic : معناي لغات و مفهوم عبارات تركيبي
- Pragmatic : سطح بالاتري از علم كه دربارة معناي متفاوت يك جمله در متون مختلف قضاوت مي كند .

- World : اطلاعات كلي كه شخص در برقراري ارتباط مي داند . شامل شناخت اهداف و عقايد ديگران نيز مي شود .

براي آشنايي با سيستم هاي «پردازش زبان طبيعي » (NLP) و شروع كار ، دانستن مفاهيم و اصول زبان و زبان شناسي لازم است . پس از اين ابزاري كه ما را در طراحي كمك و ياري مي نمايد ،گرامر است. پس از توصيف زبان به كمك گرامر ، به مرحله پياده سازي نزديك مي شويم. گام بعدي كه ما را به فهم زبان توسط ماشين نزديك تر مي سازند ، استفاده بهينه و مؤثر از Parser ها يا تجزيه گرهاست. بسته به نوع كاربرد پارسر مناسب (با توجه به گرامر تعريف شده) بايد انتخاب و پياده سازي شود .

مراحلي فوق الذكر گرچه پايه اي و مشترك هستند ولي در سيستم هاي متفاوت و براي زبانهاي مختلف ، تفاوت هاي زيادي دارند و در هر كدام الحاقات و تغييراتي مشاهده مي شود . مراحل سه گانه تحليل هاي لغوي ، نحوي و معنايي بطور كلي بر اين سيستم ها حاكم است و با اين ابزارها و خروجي هايشان (گرامر و پذيرنده و پارسر) عمل هاي خاص خود را انجام مي دهند تا درنهايت فهم مطلوب بر اساس پروژه نسبت به آن زبان طبيعي حاصل شود .

در حالت كلي برنامه هاي فهم زبان طبيعي سه روش براي انتخاب دارند[2]:

۱. استفاده از كليد واژه و تطبيق الگوها
۲. الحاق آناليز ساختاري (گرامري) و معنايي
۳. ارائه هاي نمايشنامه اي (مقايسه و تطبيق ورودي با يك وضعيت جهاني و حقيقي)

روش ۱ ساده ترين است . برنامه معروفELIZA از اين نوع است . اين روش ممكن است جملات نا مفهوم ايجاد كند. در حقيقت در اين روش ماشين زبان را نفهميده است .

روش سوم محاسبات مربوط به آناليز گرامري را حذف مي كند اما مقادير قابل توجه مشخصي از اصلاعات قبلي، مانند لغات كلي جهاني بايد داشته باشيم.

روش متداول تر ، روش دوم است. پس از دوآناليز گرامري و معنايي در اين روش، اطلاعات ساختاري از جملات ورودي بدست خواهد آمد. بدين منظور پارسرها استفاده مي شوند. مزيت اين روش قدرت آن است و نقطه ضعف آن محاسبات زياد مي باشد.

۳- پردازش زبان طبيعي فارسي

زبان فارسي از زبانهاي هند و اروپايي منشعب شده است و قدمت آن به ۱۵۰۰ سال قبل از ميلاد بر ميگردد. هرچند قدمت زبان فارسي كنوني به قرن هفتم ميلادي بر ميگردد. الفباي رايج در زبان فارسي

الفباي زبان عربي بعلاوه چهار حرف "گ پ چ ژ" است. در اين بخش ما به برخي از ويژگيهاي زبان فارسي كه از ديدگاه NLP نياز به بررسي هاي ويژه دارند مي پردازيم.

حروف در زبان فارسي به يكديگر چسبيده مي شوند تا كلمات را ايجاد نمايند. بجز حروف "ا دذررژو" بقيه حروف داراي اشكال متفاوتي مي باشند كه بسته به محل آنان در كلمه شكل متناسب نوشته مي شود. گرچه در موقع نوشتن بين كلمات فاصله گذاشته مي شود ولي اگر حروف فوق الذكر در انتهاي كلمه قرار گيرند اين احتمال وجود دارد كه فاصله بين كلمات گذاشته نشود. همچنين اگر اين حروف در وسط كلمه قرار گيرند ممكن است بعد از آنها فاصله اضافه گذاشته شود. اين موضوع تشخيص كلمات را از يكديگر مشكل مي سازد و نياز به يكسري پيش پردازشهايي جهت تجزيه جمله به كلمات مي باشد.

مسأله خاصي كه در دستور زبان فارسي مشاهده مي شود اين است كه درباره ترتيب اجزاء جمله قيد و محدوديت خاصي وجود ندارد. بدين معني كه اجزاء ساختاري جملات فارسي، مكان به هم پيوسته و ثابتي در جمله ندارند. براي روشن شدن موضوع جملات زير را در نظر بگيريد:

۱- علي ديد سيب سرخ را در مدرسه .

۲- علي در مدرسه سيب سرخ را ديد.

۳- سيب سرخ را علي در مدرسه ديد.

۴- درمدرسه ديد علي سيب سرخ را.

يك فرد بومي زبان فارسي اگرچه جمله دوم را بهتر مي پسندد اما براي درك مفهوم جمله ، با بقيه جملات مشكلي ندارد و حقيقت معنايي يكساني را از همه استنباط مي كند. چنين پديده اي درزبانهايي نظير عربي و انگليسي وجود ندارد؛ اگرجملات يكسان و درست گرامري بدين صورت در اينگونه زبانها مشاهده شود جابجايي لغات از يك جزء فراتر نمي رود و آن هم باز در محدودة تعريف شده قوانين گرامري زبان اش است . مثل افعال دو بخشي در زبان انگليسي كه بخش هايش تفكيك پذير است و مي توان مفعول را يا بين دو بخش فعل و يا پس از فعل مركب آورد.

آنچه در مثال فوق مشاهده گرديد آزادي عمل نسبتا زيادي است و به اين زبانها Free word order language مي گويند[3] . البته بايد دقت داشت كه آزادي ترتيب ، مربوط به اجزاء ساختاري مي شود يعني دو جزء ساختاري مي تواند جابجا شوند بي آنكه جمله ايراد گرامري پيدا كند واضح است اين آزادي مربوط به كلمات به تنهايي نمي شود مگر اينكه كلمه خود جزء ساختاري مستقلي باشد. البته در جملات ادبي زبان فارسي گاه جابجائي هايي به چشم مي خورد كه صرفا" مربوط به اجزاء ساختاري

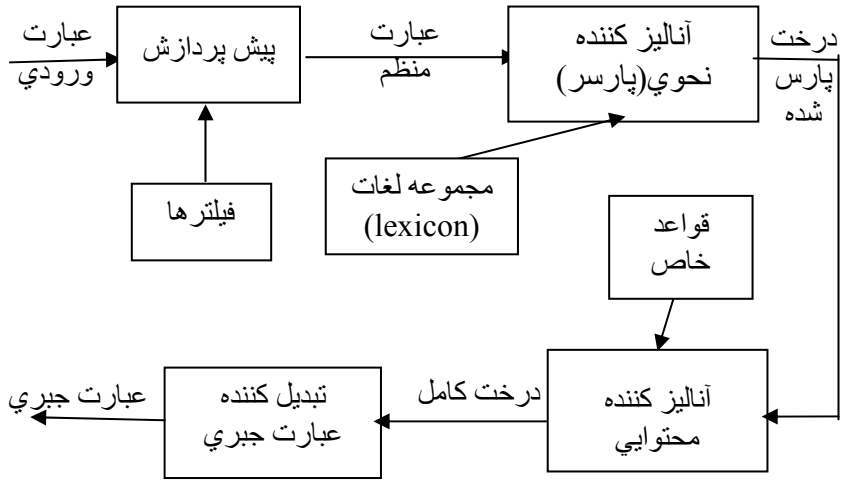
نمي باشد اما در هر حال در فارسي روان و نوشتاري كه مراد ماست خود را محدود به اجزاء ساختاري مي نمائيم.

براي زبان فارسي نيز البته بايد مشابه زبانهاي ديگر سطوح مختلف زبان شناسي از ديدگاه پردازش زبان طبيعي را بررسي كرد، اما در هر حال موارد ذكرشده توجه خاص تري مي طلبد، در هر زبان استثنائات گرامري موجود را هر چه بيشتر بتوان در قوانين لحاظ كرد، گرامر قوي تر وپيچيده ترمي شود و بازه وسيع تري از جملات زبان را مي تواند بپذيرد. وجود اين قبيل استثنائات (مثلا امكان آوردن فعل مفرد براي فاعل جمع و بي جان يا امكان آوردن فعل جمع براي فاعل مفرد و انسان از جهت احترام و....) در زبان فارسي نسبتا زياد به چشم مي خورد و كاربرد زيادشان، لحاظ كردنشان را در گرامر ضروري مي كند.

همچنين عدم وجود حروف بزرگ و كوچك (case)، عدم وجود نشانگر بين عبارات اسمي (NP)، عبارات فعلي (VP)، ... تشخيص اين عبارات را مشكلتر مي نمايد. البته كسره اضافه در بسياري از موارد عمل نشانگذاري را انجام مي دهد ولي چون نوشته نمي شود قابل بهره گيري نيست. [4]

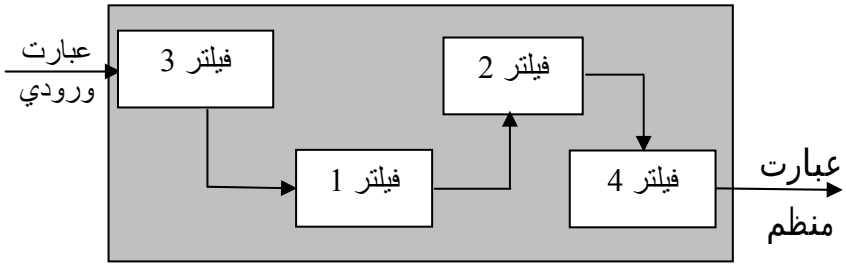
۴- ساختار سيستم

سيستمي كه در اينجا طراحي شده براي استفاده در جستجوهاي اينترنتي است كه كاربر بتواند تقاضاي جستجويش را براحتي و بي تكلف به زبان خود - زبان فارسي - بيان دارد.هدف از اين سيستم نوعي استخراج داده از جملات فارسي مي باشد و نياز به فهم كلي جمله نيست. معماري اين سيستم در شكل 1نشان داده شده است.



شكل **1**- معماري سيستم

مرحله پيش پردازش عملياتي برعبارت ورودي كه توسط كاربروارد مي شود انجام ميدهد تا جمله حالت استانداردي كه توسط بخش بعد قابل آناليز باشد پيدا كند. با توجه به اينكه فارسي يك زبان Free order از نظر ترتيب اجزاء نحوي است ، عمدة كار ثابت كردن ترتيب بعضي از اجزاي جمله است. خروجي اين مرحله كه بازهم يك جمله معقول و درست فارسي است با اعمال فيلترهاي لازم بر جملات ايجاد مي شود. ترتيب اعمال اين فيلترها در شكل2 نشان داده شده است. شماره گذاري فيلترها بر اساس ميزان اهميت و نه ترتيب اجراي آنها صورت گرفته است.



شكل **2**- ترتيب اعمال فيلترها در بخش پيش پردازش

در اين شكل هدف فيلتر۱ اين است كه فعل هر جمله كنار اجزاي خود جمله و آخر جمله باشد. هدف فيلتر ۲ اين است كه مفعول در جمله موجود باشد و نسبت به تركيبات اضافي تقدم داشته باشد. موضوع Free order بودن زبان فارسي از نظر ترتيب اجزاي جمله در اين مرحله نيز مطرح مي شود. اين فيلتر بر روي مفعول (از نوع خاص) براي گرامر مربوط كار مي كند. نقش فيلتر ۳ تطابق و تغييرواژك هاي ورودي براساس لغات lexicon است..گرچه قرار گرفتن جداكننده ها (white spaces) در ميان لغات يك جمله طبيعي است اما اگر زياد و غير ضروري باشند ممكن است در كار Scanner مشكل ايجاد كنند. اساس كار Scanner براي تشخيص token ها ، جداسازي كلمات بين جداكننده ها و مراجعه به lexicon است حال اگر در كلمات مركب فضاي خالي و Space اضافي بين اجزاء باشد، مشكل ايجاد مي شود. البته منظور اين نيست كه كلمة مركب ، اجزاء آن بي مفهوم و غير مستقل اند، بلكه اين واژه صرفا بخاطر نوع كلماتي كه در Lexicon خاص پروژه استفاده شده اند بكار رفته است. در فيلتر ۴ نيز tokenهايي كه از ديد پروژه غيرلازم اند، با توجه به lexicon ، حذف مي شوند.

تحليل نحوي و ساختاري جملات با كمك دوبخش Scannerو Parser انجام ميشود. بخش Scanner، Token هاي جمله را با خصوصيات خاصي كه در آن جمله دارند و با استفاده از lexicon تشخيص مي دهد. بخش Parser نيز با استفاده از اين token ها و قوانين گرامر طراحي شده جملات و تجزيه مي كند و در قالب درخت Parsing تحويل ميدهد.

اگر نتيجه كار حاصل از بخش آناليز نحوي مستقيما به بخش تبديل كننده عبارت جبري داده شود خروجي نهايي قابل قبول خواهد بود ، اما درصد درستي الگوريتم با استفاده از آنايز مفهومي بيشتر مي گردد. قوانين و روابط خاصي كه از نحوه بيان جملات فارسي ورودي استخراج شده ، اساس عمليات و روند كار اين بخش است.

۵- پياده سازي سيستم و نتايج

براي هر زبان طبيعي كه بخواهيم پردازشگر وسيستم درك كامپيوتري بنويسيم مسلما بايد ويژگي هاي زبان و دستورزبانش را به دقت مورد مطالعه قرار دهيم. بجز موارد بحث شده قبلي، چند نكته خاص در دستور زبان فارسي وجو دارد كه در پروژه با آن مواجه شده ايم.

كه موصولي : از موارد خاصي كه در زبان فارسي ايجاد جملات مركب و پيچيده مي نمايد، «كه موصولي» است كه بعد از اسم مي آيد و توضيحي صفت گونه پيرامون آن اسم مي دهد. چنين حالتي گرامر را پيچيده و مشكل مي كند و ما آن را با فيلتري ازحالت تودرتوبودن جملات به حالت الحاقي شدن ، تبديل كرده ايم . مثلا "اطلاعاتي در مورد موضوع X كه شامل Y نباشد، مي خواهم" به "اطلاعاتي در مورد موضوع X مي خواهم كه شامل Y نباشد." تبديل شده است. نظير اين گونه جملات در زبان فارسي بسيار است و جملات گرامري كاملا درستي هستند لذا در پروژه لحاظ شده است

جملات پايه و پيرو : جملاتي كه در فارسي به هم ملحق مي شوند دو حالت دارند:

۱- جملات مستقلي هستندكه به تنهايي مفهوم دارند.

۲- جملات وابسته اي هستند كه هدف و منظور اصلي را يكي از جملات بيان مي دارد و جملات ديگر پيرامون آن هدف و عمل اصلي توضيحاتي مي دهند.

جملات نوع دوم به جملات پايه و پيرو معروف اند. جمله اصلي ، پايه و جملات وابسته بدان پيرو هستند. بدون توجه به چگونگي تشخيص جمله پايه از پيرو ، آنچه لازم داشتيم تشخيص اين دو نوع جمله مركب از هم بود. مميزة اينها حروف ربطي است كه بينشان قرار مي گيرد و ما اين مسأله را با تفكيك حروف ربط به دو دسته در نظر گرفته ايم. درواقع اين پديده اهميت خود را در عملگر خاص حروف و در بخش تبديل كننده عبارت جبري نشان مي دهد.

بر اساس ساختارهاي سطحي مشاهده شده در جملات، گرامري خاص طراحي شده است كه ليست قوانين مربوطه به همراه اجزاء سازنده خاص آنها معرفي مي گردند. همانطور كه با مراجعه به قوانين مشاهده خواهد شد، چون سعي شده است پيچيدگي جملاتي در بخش پيش پردازش كم شود، و در

بين جملات و يا اجزاء سازنده فقط Conjunction (موصول) داريم، تشخيص و بررسيجملات ساده تر است. اين موضوع حتي جملاتي را كه بخش پايه و پيرو دارند نيز شامل مي شود.

بازه جملاتي كه گرامر تشخيص مي دهد جملات اخباري و امري مي باشد. جملات سؤالي پوشش داده نشده اند چون امكان بروز جمله سؤالي از هر نوع پيچيده كه صريحا درخواست جستجو را مطرح نكند وجود دارد. اگر جمله اي توسط گرامر پذيرفته نشود هر كدام از موارد زير احتمال دارد: ۱- اشتباه گرامري بارزي دارد. ۲- جزء جملات قابل پذيرش توسط سيستم نيست.

قوانين گرامر:

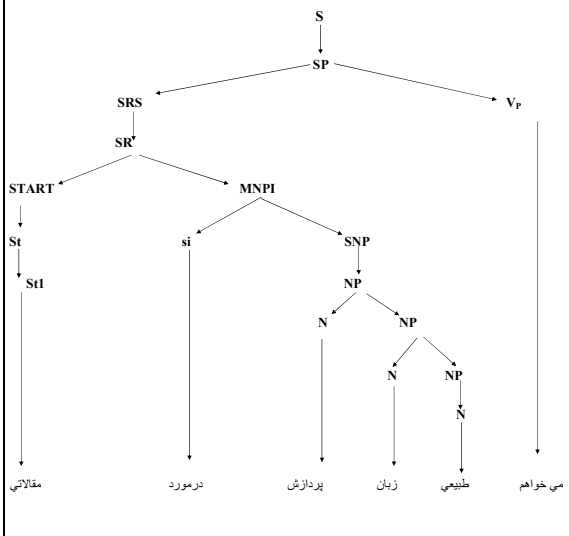
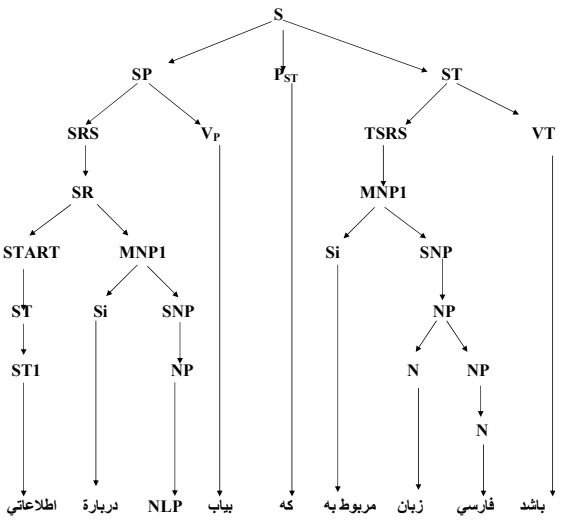
$S \rightarrow S\ P_{ss}\ S$ $S \rightarrow SP\ P_{st}\ ST \mid SP$ $ST \rightarrow ST\ P_{tt}\ ST$ $SP \rightarrow SRS\ V_p$ $SRS \rightarrow SR$ "va" $SRS \mid SR$ $SR \rightarrow$ Start MNP1 | Start MNP2 Start $\rightarrow$ St | St "ra" $MNP1 \rightarrow$ si + SNP	$MNP2 \rightarrow$ SNP "ra" $SNP \rightarrow$ NP pp SNP | NP $ST \rightarrow TSRS + V_t \mid TSRS' + V_t$ $TSRS \rightarrow$ MNP1 "va" TSRS | MNP1 $TSRS' \rightarrow$ MNP2 "va" TSRS | MNP2 $NP \rightarrow N+NP \mid N$ $St \rightarrow St1\ St \mid St1$

- معرفي غير ترمينال ها و ترمينال ها :

منظور از ترمينال ها ، علائمي هستند كه به لغات موجود در Lexicon ختم مي شوند.

غير ترمينال ها :	ترمينال ها
S : جمله يا جملات	**P_{ss}** : حروف ربط مجاز بين دوجمله غيرپايه ، پيرو (و ....)
SP : جمله پايه	
ST : جمله پيرو	**P_{st}** : حرف ربط بعد جمله پايه (كه)
Start : شروع جمله پايه	**P_{tt}** : حرف ربط بين دو جمله پيرو (و...)
SR : عبارت شامل موضوع درخواست	**St1** : نوع جستجو (كتاب ، مقاله ، site ....)
MNP : عبارت اسمي شامل موضوع جستجو	**si** : پيش عبارت اضافي مشخص كننده موضوع جستجو (درمورد و ....)
SRS: چندين SR	
SNP: موضوعات خالص Search با حروف ربط بين شان .	**pp** : حرف ربط يا اضافه
NP: عبارت اسمي	**V_p**: فعل جمله پايه
TSRS : تركيبات جستجو دار قبل از فعل در جمله پيرو	**V_t**: فعل جمله پيرو
	N: اسم (اسم خاص يا يك ورودي در فرهنگ لغات ياخارج آن)

جهت توضيحات بيشتر درخت parsing مربوط به دو مثال نشان داده شده است. در رديف آخر خروجي كه به موتور جستجو (در اينجا google)ارسال مي شود ملاحظه مي گردد

مقالاتي در مورد پردازش زبان طبيعي مي خواهم	اطلاعاتي درباره NLP بياب كه مربوط به زبان فارسي باشد
	
پردازش زبان طبيعي	NLP زبان فارسي

۶- نتيجه گيري

در اين مقاله جزئيات يك سيستم جهت پردازش جملات فارسي محاوره اي و تبديل آنان به عباراتي كه توسط موتور جستجو قابل استفاده مي باشد مورد بحث قرار گرفت. گرچه اين سيستم بعنوان نسخه اول يك سيستم كامل مي تواند تلقي گردد ولي قادر است تا يكسري جملات را تشخيص دهد.

در مورد گسترش و توسعه پروژه از ديدگاههاي خاصي مي توان اقدام كرد. مثلا ميتوان آن را تعميم داد بطوري كه براي سيستم هايي با هدف مشابه قابل استفاده باشد. در رابطه با بالا بردن كارآئي خود پروژه نيز طراحي مرحله مرحله و الگوريتمي همه قسمت هاي كارمي تواند مشخص كند عدم پذيرش جمله درست ، دقيقا از ضعف كدام مرحله ناشي شده بوده است و آن قسمت توسعه داده شود. چون خروجي هاي هر مرحله دقيقا تعريف شده است مي توان مقايسه و عدم تطبيق احتمالي اي كه در هر مرحله ايجاد اشكال كرده است را يافت . همچنين تكميل دامنه لغات سيستم به كارآيي آن كمك خواهد نمود.

۷- مراجع

1. Mohammad A. Sanamrad and Hauya Matsumoto.. *PERSIS: A Natural-Language Analyzer for Persian* . In Journal of Information Processing, 1985 وvol 8, No. 4, pages 271--279.

2. Siamak Rezaei , Mehrdad Fahimi - Toward a Machine Translation System for Persian Language - University of Edinburgh - November 23, 1994

3. *Siamak Rezaei, Matthew W. crocker "Parsing Free word order Language : Persian" :*

4. Jan W. Amtrup, Hamid Mansouri Rad,Karine Megerdoomian and Rémi Zajac - "Persian-English Machine Translation:An Overview of the Shiraz Project", New Mexico State University - April 2000.

Searching Internet Using Persian Natural Language

Mohsen Kahani (kahani@um.ac.ir)
Assistant Professor, Ferdowsi University of Mashhad.

Abstract

In this article, the use of Persian natural language for searching Internet is investigated and the result of an applicational project is explained. In the project, a Persian sentence is provided by a user, in which s/he asks for the subjects that s/he want to search internet for more information. The system processes the Persian sentence, removes the unnecessary words and extracts the keywords as a Boolean expression. The expression can then be submitted to a search engine and the result is presented to the user.

Keywords:
Persian, Natural Language Processing (NLP), Search Engine, Artificial Intelligence.