مجلـه‌ی علمی پژوهش در علوم رایانه ، شماره ۲ ، تابستان ۱۳۹۵، ص ۱۹-۷

ISSN: 2745-3673
http://www.Rcsj.ir

مجلـه‌ی علمی پژوهش در علوم رایانه (سال دوم)
شماره ۲ / تابستان ۱۳۹۵ / ص ۱۹-۷

استخراج اطلاعات از فروشگاه های الکترونیک فارسی

وحید متقی [۱]

[۱] کارشناسی ارشد فناوری اطلاعات گرایش تجارت الکترونیک،دانشگاه قم

نام و نشانی ایمیل نویسنده مسئول:

وحید متقی
mvahid500@gmail.com

چکیده

وب سایت های حراج الکترونیک و سرویس های ارائه شده توسط این سایت ها رو به افزایش است. از عامل ها در سایت های حراج الکترونیک به زبان انگلیسی استفاده زیادی می شود ولی تاکنون تحقیقی درباره طراحی عامل های هوشمند برای استخراج اطلاعات به زبان فارسی برای وب سایت های حراج الکترونیک انجام نشده است. هدف از این تحقیق طراحی عاملی است که شرایط مورد نیاز کاربر را از او دریافت کرده و کالاهایی را که مطابق با نیاز کاربر است در مدت زمان کمی به او ارائه نماید. فرآیند خرید شامل شش مرحله است برای خودکار سازی کل فرآیند خرید از سیستم های چندعاملی و برای استخراج اطلاعات از مدل حالت متناهی استفاده شده است. پس از استخراج اطلاعات عامل های نرم افزاری فرآیند خرید را تکمیل می نمایند. استفاده از این مدل باعث بالا رفتن سرعت جستجوی کالاها شده و محصولات مرتبط با نیاز کاربر را فراهم می کند..

واژگان کلیدی: حراج الکترونیک ، عامل های هوشمند، استخراج اطلاعات، زبان فارسی، مدل حالت متناهی

مجلـه‌ی علمی پژوهش در علوم رایانه ، شماره ۲ ، تابستان ۱۳۹۵، ص ۱۹-۷

ISSN: 2745-3673
http://www.Rcsj.ir

مقدمه

تجارت الکترونیک از جمله فرآیندهای جامعه اطلاعاتی دنیای امروز است که در سالهای اخیر با حضور اینترنت بسیار توسعه یافته است. در تجارت الکترونیک سه بعد اصلی تجارت ، که شامل محصول و خدمات مورد مبادله ، فرآیند فروش ، تحویل و خدمات پس از فروش می باشند ، می توانند از حالت فیزیکی و کاملا ملموس تا حالت الکترونیک تغییر نمایند. ترکیبات گوناگونی از حالت های فیزیکی و الکترونیکی ابعاد تجارت، تعیین کننده سطوح تجارت الکترونیکی بوده و در صورتیکه هر سه بعد حالت الکترونیکی داشته باشند ، بالاترین سطح در تجارت الکترونیکی شکل می گیرد. این در حالی است که در تجارت سنتی هر سه مرحله، فیزیکی و کاملا قابل لمس است و در نتیجه تجارت الکترونیکی می تواند در تمام یا بخشی از مراحل چرخه تجاری به کار گرفته شود. چرخه تجاری از جستجوی کالاها و خدمات متناسب با نیازها ، جستجوی عرضه کننده و انجام مذاکره ، سفارش، حمل و پرداخت بها، فعالیت های پس از فروش مثل گارانتی و خدمات پس از فروش تشکیل شده است.[1]

در فعالیت های خرید و فروش سنتی ، خریدار در مراحل تفسیر اطلاعات و داده های کسب شده در مورد محصولات و خدمات ، اتخاذ تصمیم بهینه خرید و در نهایت انجام مذاکره و معامله و پرداخت ، نیازمند صرف وقت و تلاش بسیار است. هدف اصلی تجارت الکترونیک حداقل کردن حضور فیزیکی و فعالیت خریدار و فروشنده در کلیه مراحل خرید و فروش و بهینه کردن این فرآیند است.[2] هوشمند سازی فرآیند های تجارت الکترونیک از کاربردهای مختلف عامل های هوشمند است. این عامل های نرم افزاری نوع فعالیت را درک کرده و فرآیند تجارت را به طور مستقل ، با دریافت اطلاعات اولیه ای از کاربر انجام می دهد. عامل های نرم افزاری هوشمند می توانند در گستره وسیعی از کاربردها از جمله : پست الکترونیکی ،حراج ها، کنترل و نظارت در تجارت الکترونیکی و تسهیل فرآیندهای طرف مشتری استفاده شوند.[3] [4]

تاکنون تحقیقات زیادی درباره کاربرد عامل های هوشمند در سایت های حراج الکترونیک انجام شده است. در یک سیستم چندعاملی خودمختار و هوشمند عامل ها نیاز دارند تا با یکدیگر کار کنند و به اهداف رایج و شخصی خودشان برسند در محیطی که منابع محدود و یا خیلی کمی دارد، اینکه چطور این منابع به طور کارا تخصیص داده شود حیاتی است. لی و مای یک دیدگاه ریاضی برای حل این مسئله ارائه داده اند.[5] حراج های آنلاین خودمختار بر اساس عامل یک موضوع تحقیقاتی در حال پیشرفت در تجارت الکترونیک است. اغلب تحقیقات بر روی مکانیزم ها و استراتژی های حراج تمرکز کرده اند اما از تحقیق بر روی بقیه موضوعات کلیدی در ساخت سیستم حراج خودمختار آنلاین مانند مکان یابی حراج و همکاری بین شرکای تجاری غفلت نموده اند. هوانگ و لیو برای حل این مسئله بر اساس تکنولوژی عامل سیار و تئوری بازی راه حلی را ارائه نموده اند. [6]

با استفاده گسترده از اینترنت حراج الکترونیک خیلی رایج شده است.اینترنت اطلاعات کاملی از بازار و زیر ساختاری برای اجرای حراج با هزینه های اجرایی کمتر را فراهم می کند. مزایده نزولی و مزایده دومین قیمت رایج ترین قالب های حراج الکترونیک است.اکایا و بدور یک مدل پویا از حراج الکترونیک ارائه داده اند برای اینکه بررسی کنند چطور رضایت خریداران به وسیله انواع مزایده های نزولی تحت تاثیر واقع می شود. این موضوع به صورت تئوری در اقتصاد درباره روشهای مختلف حراج ایستا مورد بررسی قرار گرفته است. برای غلبه بر محدودیت های این دیدگاه ، مدل جدیدی مبتنی بر عامل ارائه شده است که در آن محققان از شبیه ساز برای تحقیق رفتار و تعامل عامل های خودمختار در محیط های اقتصادی و اجتماعی استفاده می کنند.[7] در حراج خرید گروهی ، حراج کننده به دو نقش مختلف تقسیم می شود: تامین کننده و عامل حراج. رابطه عمده عامل ها بین تامین کنندگان و عامل حراج بوجود می آید.کوائن برای تحلیل این رابطه ، ابتدا جریان حراج را شرح داده و تضاد اهداف و عدم تقارن اطلاعات این رابطه عامل محور را بررسی می کند و سپس راه حلی را برای حل این مسئله ارائه می دهد.[8]

تحقیقات اخیر نشان می دهد که محیط آموزش مبتنی بر وب و یادگیری فعال می تواند کارایی یادگیری را بهبود ببخشد. چونگ سیستمی را توسعه داده است که یادگیری موثر را بوسیله یک سیستم مبتنی بر وب ایجاد نموده و مشارکت فعال فراگیران را از طریق رقابت بازی برای موضوعات تجارت الکترونیک و برنامه نویسی عامل های حراج تسهیل می کند. رقابت بازی درباره یک حراج در بین عامل های سیار برای مزایده منابع مشتریانش می باشد. عامل های سیار در واقع عامل‌های نرم افزاری هستند که به وسیله فراگیران برنامه ریزی شده اند .[9] مکانیزم حراج به طور گسترده ای در سایت های مبتنی بر وب استفاده می شود اما ممکن است در آینده زیاد کارا نباشد و نیاز به تحول عامل های حراج قابل انطباق به محیط حراج پویا باشد. چوان از برنامه نویسی شبکه ژنتیک در عامل های حراج استفاده نموده است. مدل پیشنهادی چوان به عامل ها در رشد استراتژی بوسیله خرید کالاهای بیشتر با قیمت کمتر کمک می کند. همچنین برنامه نویسی شبکه ژنتیک باعث فهمیدن استراتژی مناسب در شرایط جاری می شود.[10]

ISSN: 2745-3673
http://www.Rcsj.ir

یکی از توانایی های حیاتی عامل های هوشمند گرفتن تصمیم عاقلانه ، دقیق و سریع درون محیط پویا در یک مدت زمان منطقی است. مصباح و تقی یار یک روش دسته بندی جدید بر اساس الگوهای مثبت و منفی معرفی می کنند. برای دسته بندی از تاریخچه لاگ داده حراج TAC/AD استفاده شده است . عامل هایی که مجهز به دسته بندی کننده ها هستند می تواند سود بیشتری در مقایسه با دیگرانی که مجهز نیستند بدست آورند.[11] چن یک سیستم حراج چندکاربره و چند دسترسی را طراحی و پیاده سازی نموده است. کاربران می توانند به سیستم حراج از طریق وب ، وسایل مجهز به پروتکل کاربردی بی سیم و عامل ها دسترسی داشته باشد. سیستم حراج انواع مختلف حراج شامل حراج انگلیسی ، حراج هلندی ، حراج آمریکایی ، حراج قیمت پنهان و حراج دو طرفه را پشتیبانی می کند.[12] قانون قیمت گذاری تمایزی یا پرداخت بر اساس پیشنهاد برای جایگزینی با قانون قیمت گذاری یکنواخت در بازارهای الکترونیک ارائه شده است. با این انتظار که این مدل قیمت های بازار را کمتر و ناپایداری قیمت را کاهش می دهد. زینگ و اوکاما با استفاده از دیدگاه چند عاملی ، جایی که هر عامل انطباق پذیری قیمت های مزایده را بر اساس الگوریتم Q-learning ارائه می دهد حراج پرداخت بر اساس پیشنهاد و قیمت گذاری یکنواخت را با یکدیگر مقایسه نموده اند.نتایج تجربی نشان داده است که حراج پرداخت بر اساس پیشنهاد باعث کاهش قیمت بازار و ناپایداری قیمت شده است . [13]

وب سایت های حراج الکترونیک و سرویس های ارائه شده توسط این سایت ها رو به افزایش است. از عامل ها در سایت‌های حراج الکترونیک به زبان انگلیسی استفاده زیادی می شود ولی تاکنون تحقیقی درباره طراحی عامل های هوشمند برای استخراج اطلاعات به زبان فارسی برای وب سایت های حراج الکترونیک انجام نشده است. هدف از این تحقیق طراحی عاملی است که شرایط مورد نیاز کاربر را از او دریافت کرده و کالاهایی را که مطابق با نیاز کاربر است به او ارائه نماید. در ادامه در بخش دو به استخراج اطلاعات ، بخش سوم استخراج اطلاعات از صفحات حراج الکترونیک فارسی ، بخش چهارم معماری مدل ارائه شده و در پایان به نتیجه گیری تحقیق پرداخته شده است.

۱- استخراج اطلاعات

استخراج اطلاعات نوعی بازیابی اطلاعات (Information Retrieval) است که هدف آن استخراج اطلاعات دارای قالب از اسنادی نیمه ساختارمند یا بدون ساختار است. استخراج اطلاعات از زیر شاخه های پردازش زبان طبیعی در نظر گرفته می شود. منظور از استخراج داده های وب ، تشخیص و استخراج موارد خواسته شده از صفحات وب می باشد. همچنین امکان تجمیع اطلاعات و داده ها را از چند منبع (وب سایت ها و صفحات وب) جهت تولید خدمات ارزش افزوده از قبیل جمع آوری اطلاعات وب با توجه به نیاز مشتری ، مقایسه محصولات هنگام خرید ، جستجوهای پیشرفته و غیره را به ما می دهد.

به دلیل رشد میزان اطلاعاتی که در صفحات بدون ساختار وب وجود دارد استخراج اطلاعات وب اهمیت بسیار زیادی دارد. عامل های نرم افزاری نیاز به استخراج اطلاعات دارند تا بر اساس اطلاعات استخراج شده بتوانند روی داده های بدون ساختار پردازش انجام دهند.سیستم های استخراج اطلاعات قبلا از تکنیک های پردازش زبان طبیعی مانند گرامر و واژه نامه (lexicons and grammars) استفاده می کردند در حالی که سیستم های استخراج اطلاعات وب از روشهای یادگیری ماشین و الگو کاوی برای استخراج کردن الگوها در قالب های صفحات وب استفاده می کنند.تحلیل های زبان شناسی که برای متون غیرساخت یافته اجرا می شوند نمی توانند تگ های اچ تی ام ال یا ایکس ام ال را که در یک متن آنلاین وجود دارد استخراج کنند در نتیجه از تحلیل های زبان شناسی کمتر استفاده شده و برای استخراج اطلاعات روی وب از روشهای دیگری استفاده می شود.[14-18]

۱-۱- مسایل خط و زبان فارسی در بازیابی اطلاعات

خط و زبان فارسی مشکلاتی را برای سیستمهای ذخیره و بازیابی اطلاعات ایجاد می کند.مرتضایی به برخی از مشکلات زبان فارسی اشاره شده اند که عبارتند از: گوناگونی معادل های علمی ،ضبط اسامی ، تعیین مرز کلمات: سرهم نویسی، جدانویسی و بی فاصله نویسی ، انواع جمع ها ، صورت های مختلف نوشتاری [۱]

-گوناگونی معادل های علمی: متخصصان در بیان و انتقال یک مفهوم از اصطلاحات متفاوت استفاده می کنند .به عنوان مثال بر اساس تحقیق هاشمی برای کلمه Online ، ۱۲ معادل و برای کلمه Manual ، ۹ معادل بکار رفته است. [۲]

-ضبط اسامی: در برگردان اسامی افراد، سازمان ها، عناصر و غیره از سایر زبانها به فارسی، قاعده خاصی وجود ندارد. به عنوان مثال برای کلمه Potassium سه معادل وجود دارد: پتاسیم، پتاسیوم، پوتاسیم

-تعیین مرز کلمات: سرهم نویسی، جدانویسی و بی فاصله نویسی: شیوه خط فارسی چنان است که بسیاری از واژه ها را می توان به چند صورت نوشت. این چندگونگی شکل واژه ها، برای کامپیوتر قابل درک نیست. چرا که واژه ها را تنها به همان صورتی که ذخیره کرده

ISSN: 2745-3673
http://www.Rcsj.ir

است می شناسد و بازیابی میکند. گاه یک واژه مرکب در مکان های مختلف جدا از هم قرار می گیرد. علامت جمع ها که به صورت سرهم یا جدا نوشته شود نیز، همین وضع را در فهرست های رایانه ای ایجاد میکند. برای مثال: آب گرم کن، آب گرمکن، آبگرم کن، آبگرمکن یا علیرضا، علی رضا

-انواع جمع ها: تعدد علائم جمع(ها؛ ان؛ ات ؛ ین ؛ ون) و وجود جمع بی قاعده در زبان فارسی سبب گردیده است در پایگاه هایی که کلیدواژه ها را به صورت جمع به کار می برند، مشکلی بر مشکلات بالا افزوده شود. نمایه ساز در هنگام نمایه سازی در انتخاب بین مدارس/مدرسه ها، اساتید/استادان/ استادها، محققان/محققین و مانند آن ها، مردد است.

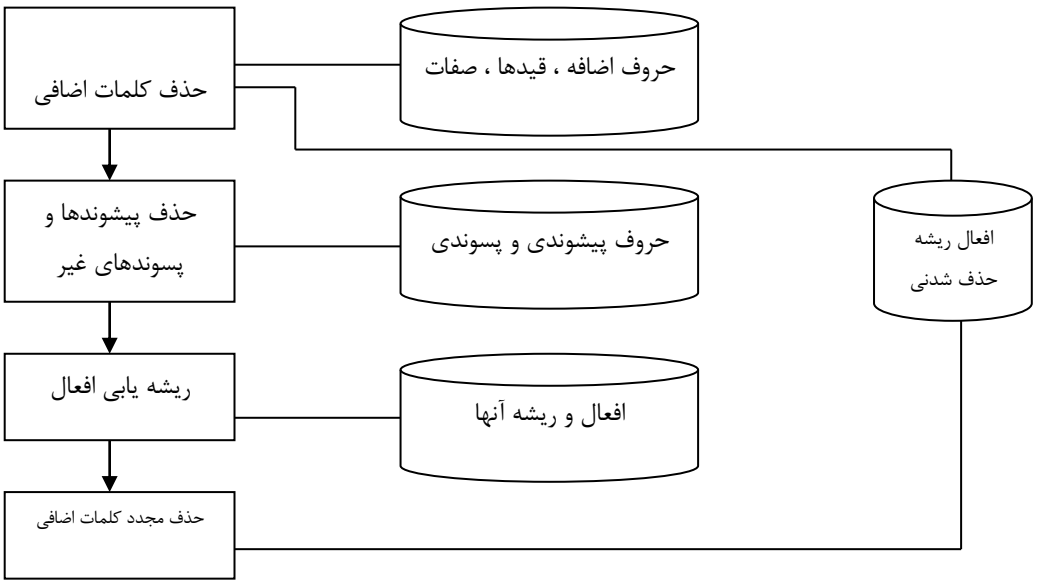
-صورتهای مختلف نوشتاری: همزه، الف مقصوره، تشدید و دوگانگی شکل نوشتاری واژه ها و اسامی، سبب ناهماهنگی‌هایی در ورود داده ها و پراکندگی اطلاعات پردازش شده می گردد. مانند هیات، هیئت و اسمعیل و اسماعیل .

۱-۲-مراحل استخراج اطلاعات

مراحل استخراج اطلاعات از صفحات وب شامل نمایه ساز، حذف کلمات اضافی ، کلمات عمومی و ریشه یابی می باشد که در زیر شرح داده می شود.

-نمایه ساز

تمام صفحات سایت حراج الکترونیک بعد از طی مراحل ذخیره سازی در مخزن، در اختیار نمایه ساز قرار می گیرد. در این بخش، اطلاعات ارسالی مورد تجزیه و تحلیل قرار می گیرد. نمایه سازی فرایند تحلیل محتوای اطلاعاتی سند به منظور استخراج کلید واژه ها به همراه ارزش آن با زبان ویژه نظام نمایه سازی است. هر عبارت مهم ای که محتویات داخل سند را تشریح کند، کلمه کلیدی گفته میشود.برای استخراج کلمات کلیدی یک سری پیش پردازش هایی باید روی متن باید انجام بگیرد که در شکل۱ مراحل کار نشان داده شده است.



شکل۱- مراحل استخراج کلمات کلیدی

-حذف کلمات اضافی

حذف کلمات اضافی نظیر حروف اضافه ، بسیاری از قیدها و صفات ، برخی از افعال (که خود ریشه اند) ، حروف ربط و غیره معمولا در مضمون کلی متن تاثیر منفی نمی گذارند، بلکه باعث خلاصه سازی متن می شود .

-کلمات عمومی

در روند نمایه سازی، واژه های عمومی در متن ورودی حذف می گردند. بعضی واژه ها در همه متون با تکرار زیاد وجود دارند. این گونه واژه ها، واژه های عمومی زبان نامیده می شوند. در واقع این واژه ها، واژه هایی مثل ضمایر، قیود، حروف اضافه و ربط هستند که در بازیابی، تاثیری بر ارزش محتوایی سند ندارد .

ISSN: 2745-3673
http://www.Rcsj.ir

-ریشه یابی

بسیاری از کلمات به کار رفته در متن، حالتهای دستوری متفاوتی از یک ریشه هستند. به منظورکاهش حجم نمایه و بالا بردن معیار بازیابی، معمولا از ریشه کلمات به جای حالت های دستوری متفاوت آنها استفاده می شود.زبان فارسی برای اشتقاق و ساخت کلمات از الحاق پسوندها و پیشوندها استفاده می کند. بنابراین ریشه یابی در زبان فارسی فرآیند حذف این الحاقات است. از طرفی متاسفانه قانون مدون و کلی برای ساخت واژگان اشتقاقی زبان فارسی وجود ندارد و در مورد هر پسوند و پیشوند استثناهای زیادی یافت می شود که کار ریشه یابی را مشکل می کند. بنابراین در مورد هر قانون باید استثنائات آن را شناسایی و نگهداری کرد، تا دقت ریشه یاب خودکار بهبود یابد.[19]

۱-۳-طبقه بندی روش های ریشه یابی

الگوریتم های گوناگون زیادی برای ریشه یابی در زبانهای مختلف از جمله زبان فارسی ارائه شده است. این الگوریتم ها را میتوانیم با توجه به نحوه عملکرد و میزان دقت آنها در دسته های جداگانه طبقه بندی کنیم. این دسته ها را در ادامه بیان می‌کنیم.

-ریشه یاب جدولی

ساده ترین روشی که برای ریشه یابی به نظر میرسد، نگهداری ریشه هر واژه در یک جدول است. در این روش با جستجوی واژه در این جدول، ریشه واژه مشخص می گردد. هر چند از این روش میتوان نتایج خوبی گرفت، اما نگهداری این جدول سربار زیادی برای سیستم خواهد داشت و تنها محدود به کلمات از پیش تعیین شده هستیم.

-ریشه یابی بر اساس الگوریتم پورتر

روش پورتر (Porter) یک روش توانمند و در عین حال یکی از قدیمی ترین روش های ریشه یابی در زبان انگلیسی است.در این الگوریتم ، برنامه از درون ساختاری تصمیم گیرنده مانند یک فلوچارت عبور کرده و با افزودن و کاستن وندها با رعایت قواعد املایی و دستوری، سعی در یافتن ریشه کلمات یا بطور خاص افعال دارد. این الگوریتم برای زبان انگلیسی طراحی شده است. مشکل این الگوریتم برای کلمات جدا از هم است. با توجه به اینکه در فارسی مرز دقیق کلمات مشخص نیست، برای کلمات چندپاره این روشها خوب عمل نمی کنند.مگردومیان یک سیستم ساده مبتنی بر قانون برای افعال فارسی ارایه شده است[20]

-ریشه یابی بر اساس مدل حالت متناهی((Deterministic finite-state machine(DFA)

ریخت شناسی بر اساس مدل حالت متناهی، روش متداولی است که در تحقیقات دفتری نژاد و مگردومیان نمونه ای از آن را میتوان دید[21] [۳] اساس کار آنها بر یک مدل زبان جهانی است که در تحقیق بیسلی ارایه شده است. [22] تعریف الگو بر اساس عبارات منظم انجام می شود و پیاده سازی آن بر اساس مدل ماشین حالت-متناهی است. طراحی پردازش گر ریخت شناسی حالت متناهی را میتوان به دو بخش مجزا تقسیم کرد: بخش مربوط به طرح زبانی و بخش مربوط به طرح رایانه ای [23].منظور از طرح زبانی، ارایه توصیف نظری کامل از ریخت شناسی افعال در زبان فارسی است. ارایه توصیف جامع از ریخت شناسی در زبان فارسی به گونه ای که قابل کاربرد در برنامه های رایانه ای باشد، نخستین گام جهت طراحی برنامه های کاربردی است. این توصیف می بایست تمام صورت های تصریفی فعل در زبان فارسی را ارایه دهد. بخش دوم، طرح رایانه ای است. در این بخش نیازهای سخت افزاری و نرم افزاری پردازشگر تعریف میشود، طرح زبانی پیاده سازی شده و ویژگیها و ساختار داخلی فایلهای برنامه تشریح میگردد.

-ریشه یابی به کمک روش های آماری

در این روش مجموعه ی بزرگ از کلمه ها با ساخت های گوناگون گردآوری میشود. هرچه این مجموعه بزرگتر و کاملتر باشد این ریشه یاب ها بهتر کار میکنند. در این روش تحلیل آماری به کار گرفته می شود. با روش آماری وندهایی که در کلمه ها تکرار شده اند، شناسایی میگردند. این روش به زبان بستگی ندارد و این بزرگترین برتری این روش میباشد. البته این روش با سه مشکل بزرگ روبروست: ۱- در این روش به مجموعه ی بزرگ از کلمه ها نیاز است. این مجموعه باید کامل باشد و کلمات درون آن نیز درست باشند. جمع آوری مجموعه ی بزرگی از کلمات صد در صد درست فارسی نیز، غیر ممکن است. ۲- در این روش نیاز به کامپیوترهایی با سرعت و حافظ زیاد است و اجرای برنامه های نوشته شده بر پایه ی این روشها بسیار زمانبر است.

۱-۳-۱- تحقیقات انجام شده درباره ریشه یابی فارسی

از جمله کارهای انجام شده در زمینه ریشه یابی کلمات فارسی میتوان به پروژه بن[23] ، ریشه یاب آماری [۴][24] اشاره نمود. پروژه بن یک ریشه یاب خاص زبان فارسی است که به عنوان جزئی از یک موتور بازیابی مورد استفاده قرار می گیرد. الگوریتم این ریشه یاب شبیه ریشه یاب پورتر است. اولین قدم الگوریتم پیدا کردن زیر رشته ای از لغت ورودی است که در لیست پسوندهای فارسی(که از روی

ISSN: 2745-3673
http://www.Rcsj.ir

گرامر فارسی تهیه شده است) وجود داشته باشد. اگر بیشتر از یک پسوند برای لغت پیدا شد، الگوریتم طولانی ترین پسوندی را انتخاب میکند که تعداد حروف ریشه (بخش اصلی لغت) را کمتر از حد مجاز نکند. (مثلا در اینجا کمترین تعداد حروف برای ریشه ۳ کاراکتر است) مثلاً برای لغت دستشان میتوان دو پسوند ان و شان را دید که شان طولانیتر است و چون حروف باقی مانده دست ، ۳ حرف یا بیشتر هستند،مشکلی برای انتخابش وجود ندارد. در این کار برای تعیین پسوند آخر لغت از یک DFA استفاده شده است که ورودی آن وارون شده ی رشته ی (کلمه ی) ورودی است و همه حالتها در آن حالت نهایی اند.

بن یک ریشه یاب حذف وند است. یعنی در هر قدم پسوندها یا پیشوندهایی را برمی دارد تا به لغت اصلی برسد. دیکشنری بن شامل مصدر و بن مضارع فعل هاست. الگوریتم بن به این صورت است که بیشترین کاراکترهای ممکن را از لغت برمی دارد (برمبنای قواعدی) و این کار را آنقدر تکرار میکند تا دیگر امکانپذیر نباشد. ولی با این روش ریشه ی به دست آمده ممکن است صحیح نباشد. مثلا با برداشتن پسوند ی از لغت خانگی ، ریشه ی خانگ به دست میآید. برای حل این مشکل، بن از روش Recoding استفاده میکند که تبدیلی به شکل AXC->AYC است و در آن C و A زمینه تبدیل را مشخص میکنند و x رشته ی ورودی و Y رشته تغییر یافته است.

ریشه یاب طراحی شده در نمایه ساز سینا مشابه ریشه یاب پورتر برای زبان انگلیسی است .[25] هر دو ریشه یاب کلمه را با یک سری پیشوندها و پسوندها در چند مرحله تطابق میدهند تا پسوندها و پیشوندها حذف شوند و ریشه کلمه به دست آید. تفاوت این ریشه یاب ها به تفاوت زبان آنها برمیگردد. الگوریتم پورتر الگوهایی از حروف صدادار و بیصدا برای تخمین محتوای اطلاعات مشخص می کند. در فارسی بسیاری از حروف صدادار نوشته نمی شوند. لذا ریشه یاب نمیتواند از آنها استفاده کند. در این ریشه یاب برای رفع این مشکل از روش تعریف حداقل طول ریشه استفاده نموده است . تفاوت دیگر این ریشه یاب با ریشه یاب پورتر در تشخیص پیشوند است، ریشه یاب میتواند
پیشوندها را مشخص کند در حالیکه ریشه یاب پورتر الگوریتمی برای تشخیص پیشوند ارائه نداده است.

۲-استخراج اطلاعات از فروشگاه های الکترونیک فارسی

همان طور که در بخش قبل شرح داده شده اولین مرحله برای استخراج اطلاعات نمایه سازی می باشد که شامل حذف کلمات اضافی و عمومی و در مرحله بعدی ریشه یابی می باشد.در ادامه هر یک از این مراحل برای زبان فارسی انجام می گیرد

۲-۱-حذف کلمات اضافی و کلمات عمومی

کلمات عمومی و کلمات اضافی را در یک فایل متنی وارد نموده ، در هنگام اجرای برنامه این کلمات از فایل خوانده شده و در ساختمان جدول درهم سازی ریخته می شود.هنگام بررسی هر مستند فارسی،کلمات اضافی موجود در این صفحه با جستجو در جدول درهم سازی ، شناخته شده و از مستند حذف می شوند.

۲-۲-ریشه یابی فعل ها بر اساس مدل حالت متناهی

طراحی پردازش گر ریخت شناسی حالت متناهی را میتوان به دو بخش مجزا تقسیم کرد: بخش مربوط به طرح زبانی و بخش مربوط به طرح رایانه ای.

-طرح زبانی: در فعل مفهوم شخص وجود دارد یعنی گوینده فعل را به خود نسبت دهد یا به شخص دیگری که به انواع زیر تقسیم می شود: اول شخص یا متکلم،دوم شخص یا شنونده و سوم شخص .در فعل مفهوم مفرد یا جمع بودن وجود دارد و فعل زمان را نیز نشان می دهد.

بن مضارع: ساخت های زمان حال (مضارع) و امر از آن ساخته می شود. جدول ۱ ساخت های زمان حال و امر را نشان می دهد.

مجلـه‌ی علمی پژوهش در علوم رایانه ، شماره ۲ ، تابستان ۱۳۹۵، ص ۱۹-۷

ISSN: 2745-3673
http://www.Rcsj.ir

جدول ۱- ساخت زمان حال و امر

ساخت های زمان حال و امر		
اجزائی که پیش از بن می آید	بن مضارع	اجزائی که پس از بن می آید
می	خور	م
می	خور	ی
می	خور	د
ب	خور	ید
ب	خور	ند

بن ماضی: ساخت های زمان گذشته و آینده از آن ساخته می شود.جدول ۲ ساخت هایی را که از بن ماضی ساخته می شود، نمایش می دهد.[۵]

جدول ۲- ساخت زمان گذشته و مستقبل

ساخت های زمان های گذشته و مستقبل		
اجزائی که پیش از بن می آید	بن ماضی	اجزائی که پس از بن می آید
	خورد	م
	خورد	ه ام
می	خورد	ه ای
می	خورد	م
	خورد	ه یودیم
	خورد	ه یاشید
	خورد	ه یوده ایم
خواهی	خورد	

۱-۲-۲-ماشین تعیین پذیر حالات متناهی تشخیص فعل های مضارع

- بخش مربوط به طرح رایانه ای : فعل مضارع شامل انواع زیر است:

ساده: (روم- روی- رود- رویم- روید- روند)

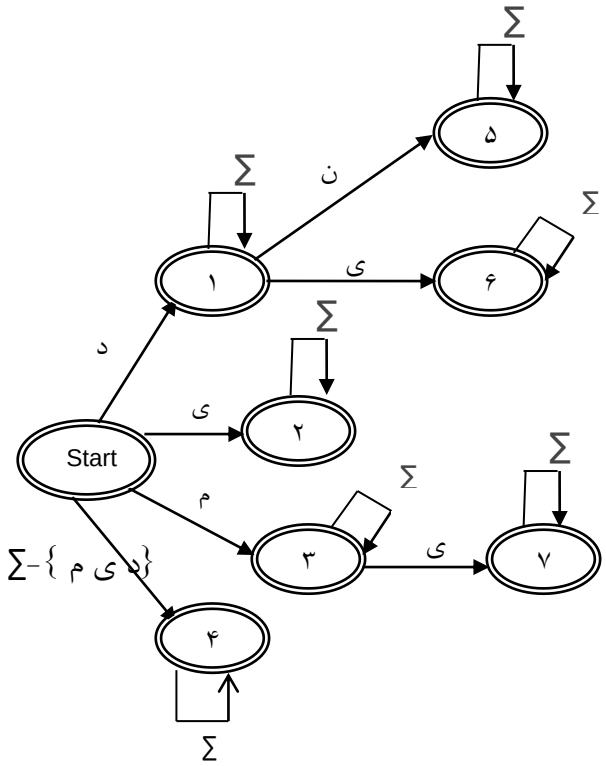
اِخباری: (می‌روم- می‌روی- می‌رود- می‌رویم- می‌روید- می‌روند)

التزامی: (بروم- بروی- برود- برویم- بروید- بروند)

ملموس: (دارم می‌روم- داری می‌روی- دارد می‌رود- داریم می‌رویم- دارید می‌رویم- دارند می روند)

الف)تشخیص پسوند: شکل ۲ ماشین تعیین پذیر حالات متناهی برای تشخیص پسوند را نمایش می دهد.

ISSN: 2745-3673
http://www.Rcsj.ir

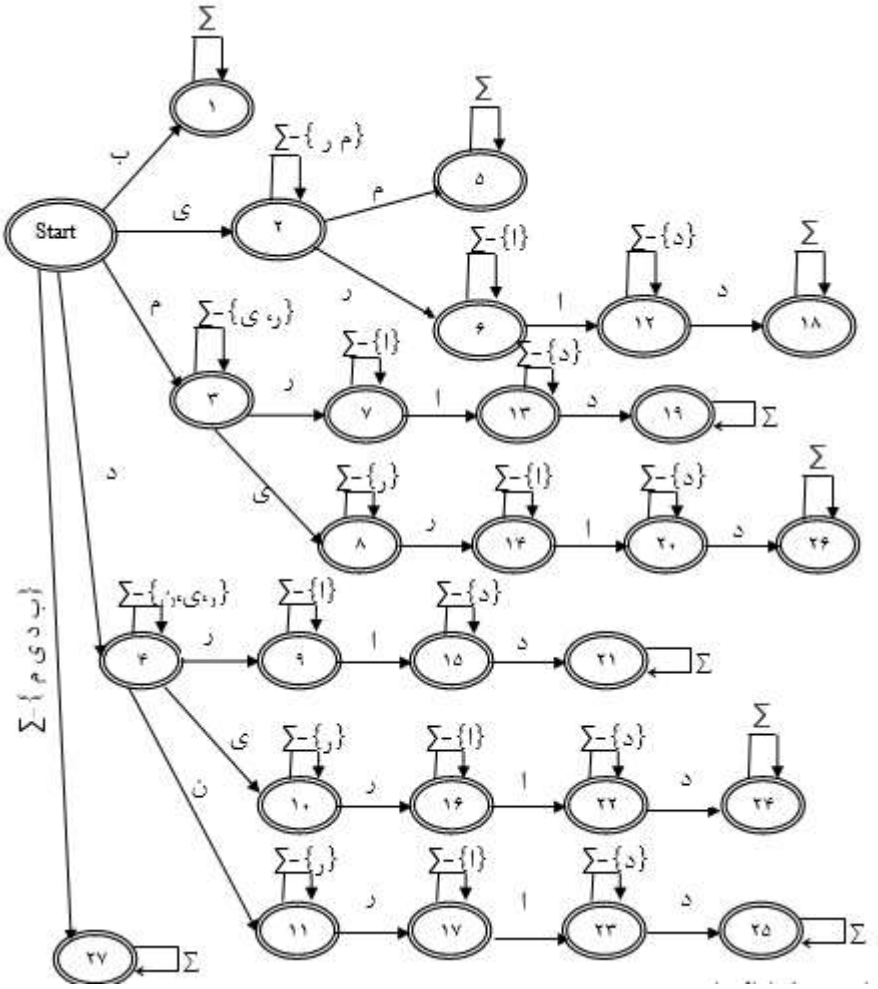


شکل ۲- ماشین تعیین پذیر حالات متناهی تشخیص فعل های مضارع بخش پسوند

پس از تشخیص پسوندهای فعلی ، زیر رشته ورودی برای تشخیص پیشوندهای فعلی به ماشین‌های تعیین‌پذیر حالات متناهی زیر ارسال می شود.

ب) تشخیص پیشوند: شکل ۳ ماشین تعیین پذیر حالات متناهی برای تشخیص پیشوند را نمایش می دهد.

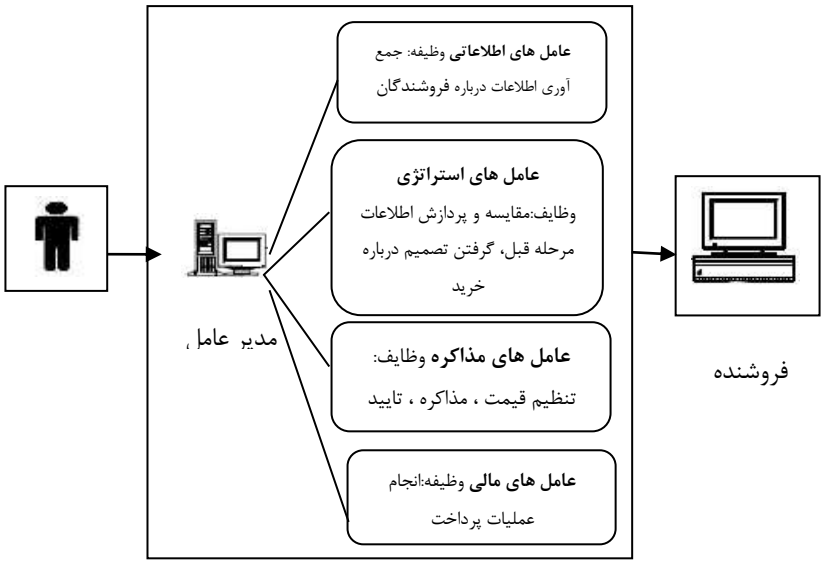
ISSN: 2745-3673
http://www.Rcsj.ir



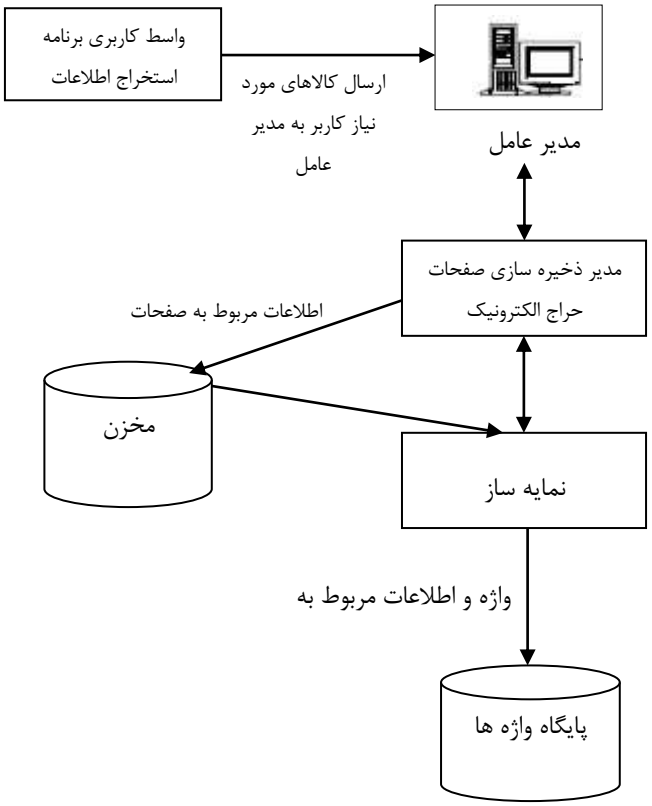
۳- معماری مدل ارائه شده

معماری مدل ارائه شده دارای دو بخش می باشد. بخش اول مربوط به مدیریت عامل های نرم افزاری و بخش دوم مربوط به فرآیند نمایه سازی سایت های حراج الکترونیک است.فرآیند خرید که توسط عامل های نرم افزاری انجام می شود، شش مرحله دارد:۱- مرحله تشخیص نیاز ۲- جمع‌آوری اطلاعات درباره محصولات ۳- ارزیابی فروشندگان ۴- مذاکره ۵- پرداخت و تحویل ۶- خدمات پس از فروش و ارزیابی . روش جستجوی کالا روی وب و مذاکره برای بدست آوردن بهترین قیمت در تحقیقات زیادی مورد توجه قرار گرفته است [26-30].در این تحقیقات اغلب از تکنولوژی عامل ها استفاده شده است به دلیل اینکه در فرآیند خودکار سازی نیاز به هوشمندی می باشد. پرداخت آخرین مرحله از یک تراکنش مالی می باشد. بدون خودکار سازی پرداخت کل فرآیند اتوماتیک نمی شود. مدیر پرداخت از تجمع سیستم های چند عاملی ایجاد شده است. عامل های اطلاعاتی، استراتژی و مذاکره فعالیت های قبل از پرداخت را انجام می دهند. این عامل ها منتظر ارسال وظایف از طرف مدیر عامل می باشند. پس از پایان مرحله خرید ، مدیر عامل ، عامل مالی را فراخوانی می کند تا عملیات مربوط به پرداخت را انجام دهد. مدیر عامل یک عامل نرم افزاری است که از طرف کاربر امور هماهنگی عامل ها را به عهده دارد.به وسیله مدیر عامل، کاربر عملیات خودکار سازی خرید کالا را طبق شرایط خود تنظیم می کند.شکل۴ ساختار چند عاملی مدیر پرداخت را نمایش می دهد. پس از اینکه عامل مذاکره پایان عملیات خرید را اعلام کرد، مدیر عامل به عامل مالی شروع عملیات پرداخت را اعلام می کند.

ISSN: 2745-3673
http://www.Rcsj.ir



شکل ۴- معماری چند عاملی مدیر پرداخت



شکل۵- معماری کلی مدل ارائه شده

ISSN: 2745-3673
http://www.Rcsj.ir

۱-۳- ارزیابی مدل

برای ارزیابی مدل ارائه شده ده کاربر به صورت عادی به جستجوی کالاهای مورد نیاز خود بر روی سایت های حراج الکترونیک پرداختند. درخواست های این ده کاربر نیز به طور همزمان به وسیله عامل های نرم افزاری انجام شد. نتایج بدست آمده توسط عامل های نرم افزاری به نیاز کاربران نزدیکتر بوده و از طرفی در زمان بسیار کمتری بدست آمد.

۴- نتیجه گیری

توسعه سیستم های رایانه ای و گسترش استفاده از فناوری اطلاعات در زندگی روزمره باعث شده تا اطلاعات از درجه اهمیتی بالایی برخوردار شوند، چنانکه عصر حاضر را عصر اطلاعات نامیده اند.میزان اطلاعات تولید شده و میزان استفاده از اطلاعات ، دو معیار اساسی برای توسعه کشورها به شمار می آیند. هر چه حجم اطلاعات افزایش می یابد کنترل و مدیریت آن مشکلتر می شود لذا تولید و وجود اطلاعات به تنهایی کافی نیست بلکه باید ابزارهایی برای استفاده از این اطلاعات فراهم شوند. در واقع کاربران باید بدانند که چگونه باید به نیاز اطلاعاتی خود در این حجم عظیم منابع اطلاعاتی پاسخ دهند.درنتیجه روشهای استخراج اطلاعات در قالب پاسخ دهی به نیاز اطلاعاتی کاربران اهمیت ویژه ای پیدا میکند. وب سایت های حراج الکترونیک و سرویس های ارائه شده توسط این سایت ها رو به افزایش است. از عامل ها در سایت های حراج الکترونیک به زبان انگلیسی استفاده زیادی می شود ولی تاکنون تحقیقی درباره طراحی عامل های هوشمند برای استخراج اطلاعات به زبان فارسی برای وب سایت های حراج الکترونیک انجام نشده است. در این مقاله مدلی ارائه شده است که شرایط مورد نیاز کاربر را از او دریافت کرده و کالاهایی را که مطابق با نیاز کاربر است جستجو می نماید.استفاده از این مدل باعث بالا رفتن سرعت جستجو شده و کالاهای مرتبط با نیاز کاربر را فراهم می کند

ISSN: 2745-3673
http://www.Rcsj.ir

منابع و مراجع

۱. ل. مرتضایی، "مسایل خط و زبان فارسی در ذخیره سازی و بازیابی اطلاعات" ، فصلنامه اطلاع رسانی ، دوره ۱۷ ، صفحه ۱۰-۱۵، ۱۳۸۰

۲. ا.هاشمی، واژگان کتابداری و اطلاع رسانی، تهران: دبیرخانه هیئت امنای کتابخانه های کشور، ۱۳۷۶

۳. ا. دفتری نژاد، "ساختواژه حالت-متناهی: روشی مناسب برای طراحی پردازشگر ساختواژی"، در هفتمین همایش زبانشناسی ایران،۱۳۸۶

۴. م.نصیری ،م .ش اسماعیلی وک. ابولحسنی، "یک ریشه یاب آماری برای زبان فارسی"، در مجموعه مقالات یازدهمین کنفرانس بین المللی کامپیوتر ،۱۳۸۴

۵. حسن انوری و حسن احمدی گیو ، ، دستور زبان فارسی ، تهران: فاطمی ، چاپ دوازدهم،۱۳۷۴

6. Rimmel G, Clement.M and Runte, M,"Intelligent Software Agents Implication For Marketing In Ecommerce", Springer Verlag, pp. 19- 33, 2000

7. Dzeng R.J and Chun Lin Y, "Intelligent agents for supporting construction procurement negotiation", Computer Law & Security Report Vol. 20 no. 1, 20-27, 2004

8. Pivk.A and Gams.M, "E-commerce Intelligent Agents", Communications of the ACM, Vol. 42, No. 3, 79- 80,2009

9. Kowalczyk.R, Ulieru.M and Unland, R, [2013], Integrating Mobile and Intelligent Agents in Advanced e-Commerce: A Survey[online], Available from: http://www.old.netobjectdays.org

10. Li.B and Ma.Ya," An Auction-based Negotiation Model in Intelligent Multi-agent System", International Conference on Neural Networks and Brain, 2005

11. Huang. J, liu.D and Yang.B, "Online autonomous auction model based on agent", Proceedings of International Conference on Machine Learning and Cybernetics, , 2004

12. Akkaya.B and Darcan.O, "A study on internet auctions using agent based modeling approach", International Conference on Management of Engineering & Technology, 2009

13. Qian.D, "The principal-agent relationships in group buying auction",2nd International Conference on Management Science and Electronic Commerce (AIMSEC), 2011

14. Cheung.R, Wan.C and Cheng.C,"an Active Learning System for Auction Agent Programming with Competitions", 9th International Conference on Computer and Information Science (ICIS), 2010

15. Yue.C ,Mabu.S and Chen.Y,"Agent bidding strategy of multiple round English Auction based on genetic network programming",ICCAS-SICE, 2009

16. Mesbah.S and Taghiyareh.F, "A new sequential classification to assist Ad auction agent in making decisions", 5th International Symposium on Telecommunications (IST), 2010

17. Chan.H, Ho.I.S.K. And Lee.R, " Design and implementation of a mobile agent-based auction system",IEEE Pacific Rim Conference on Computers and signal Processing,ACRIM, 2001

18. Xiong.G, Okuma.S and Fujita.H, " Multi-agent based experiments on uniform price and pay-as-bid electricity auction markets", Proceedings of the International Conference on Restructuring and Power Electric Utility Deregulation Technologies, 2004

ISSN: 2745-3673
http://www.Rcsj.ir

19. Yunfei. Gong, “Automatic web page segmentation and information extraction using conditional random fields”, in 16th International Conference on Computer Supported Cooperative Work in Design, 2012

20. Jellouli.I,” an ontology-based approach for web information extraction”, Colloquium in Information Science and Technology (CIST), 2010

21. Hao. Han, “a Method for Integration of Web Applications Based on Information Extraction”, Eighth International Conference on Web Engineering, 2008

22. Liu.Li,”Web Information Extraction Algorithm Based on Ontology and DOM Tree”, International Conference on Computational Intelligence and Software Engineering, 2010

23. Xinchao. Han, “Research on Web information extraction based on spider algorithm and DOM thinking”, International Conference on Information Networking and Automation (ICINA), 2010

24. Manning. C, Raghavan.D and Schütze.H, Introduction to Information Retrieval: Cambridge University Press, 2008

25. Megerdoomian.K, and Karine.A, “ Developing a Persian Part-of-Speech Tagger”, in First Workshop on Persian Language and Computers, Tehran University, Iran, 2004

26. Megerdoomian.K,”Finite-State Morphological Analysis of Persian”, in Workshop on Computational Approach to Arabic Script-Based Languages, 2004

27. Beesley.K, and Karttunen.L, Finite State Morphology: Stanford, CSLI, Publications, 2003

28. Tashakori. M, Meybodi.M and Oroumchian. F, “Bon: The Persian stemmer”,Lecture Notes in Computer Science (LNCS), Springer Verlag, vol. 2510, pp. 487-494, , 2002

29. Taghva.K, Beckley.R and Sadeh.M, “ A Stemming Algorithm for the Farsi Language”, in Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05) - Volume I - Volume 01, 2005

30. Porter M. F, “an algorithm for suffix stripping”, Program 14(3), pp. 130-167, 1980

31. Xin Wang, Shen Georganas, “ A Fuzzy Logic Based Intelligent Negotiation Agent (FINA) in Ecommerce”, Canadian Conference on Electrical and Computer Engineering,Ottawa, Canada, 2006

32. Lasheng Yu Masabo and Lian Tan.E, “ Multi-Agent Automated Intelligent Shopping System (MAISS)”, the 9th International Conference for Young Computer Scientists, Hunan, 2008

33. Vartic.R and Letia,” Rules for Representing and Handling Contracts”,IEEE International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, 2007

34. Jiang Weijin and Yao Lina, “Research on MAS behavior and paradigm learning-based evolutionary method and its application in E-commerce”, International Symposium on Computer Communication Control and Automation, Tainan, 2010

35. Cairo. O and Olarte J.G, “A negotiation strategy for electronic trade using intelligent agents”, International Conference on Web Intelligencessed , November 2004