

بسمه تعالی

وزارت علوم، تحقیقات و فناوری
مرکز منطقه ای اطلاع رسانی علوم و فناوری

عملکرد گروه پژوهشی زبانشناسی رایانه ای
سال ۱۳۸۹

فروردین ۱۳۹۰

فهرست مندرجات

صفحه	عنوان
۱	مقدمه
۸	اهداف مهم گروه پژوهشی
۱۰	طرحهای تحقیقاتی پایان یافته
۱۲	طرحهای تحقیقاتی در دست اجرا
۱۷	مقالات چاپ شده
۲۴	سخنرانی های انجام شده

مقدمه:

زبان‌شناسی رایانه‌ای یا زبان‌شناسی محاسباتی (Computational Linguistics) حوزه‌ای میان رشته‌ای است که سعی دارد با بهره‌گیری از روش‌های آماری و مبتنی بر قاعده (rule – based)، از منظر محاسباتی و مدل‌سازی زبان طبیعی بپردازد. به طور کلی، زبان‌شناسی محاسباتی از همکاری دانشمندان و کارشناسان رشته‌های زبان‌شناسی، علوم رایانه‌ای، هوش مصنوعی، ریاضی، منطق، علوم شناختی، روان‌شناسی شناختی، روان – زبان‌شناسی، مردم‌شناسی، عصب‌شناسی و برخی دیگر از رشته‌ها استفاده می‌کند.

رایانه اگرچه از دستاوردهای غرب است اما در دو دهه اخیر آنچنان در جوامع مختلف از جمله جامعه ما رخنه کرده که امروزه در بعضی از مراکز حتی تصور اینکه روزی رایانه در آن مرکز نباشد، بسیار مشکل است. با وجود این هنوز تنها تعداد محدودی از زبانها با رایانه تطبیق داده شده و در آن قابل استفاده می‌باشند و زبان فارسی از این جهت در ابتدای راه می‌باشد.

زبان‌شناسی رایانه‌ای به قسمتی از زبان‌شناسی گفته می‌شود که مسائل مربوط به زبان و رایانه را بطور عمومی مورد بحث و بررسی قرار می‌دهد و نشان می‌دهد که زبان‌شناسی چگونه می‌تواند به حل مسائل موجود در مورد تطبیق خط و زبان در رایانه، کمک نماید.

اعضاء این گروه سعی دارند که با اجرای طرحها و نشر کتابها و مقالات و سخنرانیها و ... بتوانند در برطرف کردن مشکلات موجود در خصوص هم‌آهنگی خط

و زبان فارسی با رایانه کمک کرده و در ایجاد یک نگرش علمی نسبت به آن موثر واقع شوند.

در زیر به تعدادی از موارد کاربرد علم زبانشناسی رایانه ای اشاره می شود:

۱-۱- ساخت ریشه ساز زبان فارسی: ریشه سازی یکی از ابزارهایی است که در بازیابی اطلاعات در برخورد با مسئله عدم انطباق واژگان مورد استفاده قرار می گیرد. (یعنی عدم تطبیق واژه های پرس و جو با واژه های مدرک) به فرایند حذف هر پسوند از واژه ها و تبدیل این واژه ها به ریشه های آنها ریشه سازی گفته می شود. برای مثال ریشه سازی واژه "ورزش" را به شکل "ورز" (ریشه مضارع فعل ورزیدن) به وجود می آورد. ریشه سازها عموماً متناسب با هر زبان خاص تهیه می شوند. طراحی ریشه سازها مستلزم خبرگی زبانشناسی در زبان و درک نیازهای مرتبط با بازیابی اطلاعات می باشد.

۱-۲- فرایافت Concept و تجزیه زبانهای برنامه نویسی

نوآم چامسکی توانست تشابه زبانهای طبیعی و زبانهای برنامه نویسی را به اثبات برساند. یعنی یک زبان رایانه ای مانند زبانهای طبیعی دارای یک دستور زبان و یک فرهنگ می باشد. تفسیر یک متن از تجزیه واژه ها lexicon آغاز سپس با تجزیه نحو syntax و در آخر با تجزیه مفهوم semantic آن پایان میابد.

۱-۳- ترجمه ماشینی

این شاخه از زبان شناسی رایانه ای زمان درازی کم اهمیت جلوه می کرد. اما امروزه یکی از موارد مورد علاقه پژوهشگران این رشته می باشد. پس از مرحله تجزیه واژه ها و نحو می باید تجزیه مفهوم و سپس پرگماتیک را نیز افزود. در

واقع این دو سعی در شناخت مفهوم خاص یک واژه در مکانی که ظاهر می شود را دارد.

۱-۴- پرسش و پاسخ با زبانهای طبیعی

این ایده مدتی به عنوان پاسخی قانع کننده به مسئله ارتباط انسان و ماشین تلقی می شد. این دید در واقع جنبه وسیعتری از دستور زایشی چامسکی است.

۱-۵- صرف محاسباتی

به مطالعات مربوط به ساختارهای درونی کلمات صرف گفته می شود. اغلب دست آوردها و نتایج تحقیقات در صرف محاسباتی در سایه تلاشهای علمی انسان به منظور ایجاد و ساخت سیستم های پردازش زبانهای طبیعی انسانی فراهم آمده است.

۲- ساخت ماشین های ترجمه: ترجمه ماشینی یکی از نخستین استفاده

هایی است که از پردازش زبان طبیعی به عمل آمده است. برای ساخت اینگونه ماشینها استفاده از تئوریهای نحوی و معنایی موجود در زبانشناسی ضروری است. در این زمینه زبانشناسان می توانند در حیطه های: ایجاد سیستم های ترجمه تماماً خودکار، ترجمه ماشینی نیازمند به انسان، ترجمه با کمک ماشین، ساخت ویرایش-گرهای املائی، ایجاد سیستم های ترجمه روی خط، ایجاد اصطلاحنامه های روی خط و سرانجام ماشین های همگرا و واگرا مفید باشند.

نکته حائز اهمیت آنکه برای ساخت ماشین های ترجمه از نظریه اطلاعات و رمزنگاری نیز استفاده می شود که این ارتباط حیطه های اطلاع رسانی با زبانشناسی رایانه ای را نشان می دهد. بدیهی است مراکز اطلاع رسانی برخوردار

از این مزیت قادر خواهند بود اطلاعات مورد نیاز کاربران را به بیش از یک زبان
ارایه نمایند.

۲-۲- ترجمه ماشینی (Machine Translation – MT) زیر شاخه ای

از زبان شناسی محاسباتی می باشد که عبارت است از ترجمه متنی از یک زبان
طبیعی به زبانی دیگر، توسط کامپیوتر. در سطح مقدماتی، ترجمه ماشینی یک
جایگزینی ساده برای کلمات از زبان طبیعی به زبان دیگری است. با استفاده از
تکنیک های زبان شناسی پیکره ای، ترجمه های پیچیده بیشتری قابل دستیابی
هستند. همچنین این تکنیک ها کنترل بهتر تفاوت های گونه شناسی در زبان،
تشخیص عبارات و ترجمه اصطلاحات را به خوبی و درستی جدا کردن عبارات
نامتعارف در متن، مقدور می سازند.

نرم افزارهای ترجمه ماشینی کنونی اغلب به کاربر اجازه تغییر دلخواه براساس
حوزه کاری یا حرفه ای دلخواه را می دهد (مانند: گزارش آب و هوا) در واقع ارتقاء
کیفیت خروجی با استفاده از محدود کردن کلمات جایگزین شونده، انجام می شود.
این تکنیک بطور خاص در حوزه رسمی با زبانهای فرموله شده استفاده می شود. این
بدین معنی است که ترجمه ماشینی از اسناد قانونی و دولتی آسانتر از تولید خروجی
قابل استفاده از مکالمات یا متون غیر چهارچوب بندی شده دیگر است. همچنین
کیفیت خروجی بهبود یافته می تواند با استفاده از دخالت انسان بدست آید. برای
مثال سیستم هایی موجودند که اگر کاربر بطور کاملا واضحی کلماتی که اسامی
خاص هستند را معین کرده باشد، قادر به ترجمه دقیقتری هستند. با کمک گرفتن از
این تکنیک ها ترجمه ماشینی بعنوان یک ابزار برای کمک کردن به مترجمان

(انسانها) و بسیاری از موضوعهای محدود، قادر به تولید خروجی قابل استفاده و نهایی است.

در ترجمه ماشینی ویژگیهایی وجود دارد که نه تنها از نظر جاذبه و کشش علمی، بلکه از دیدگاه اقتصادی و دیگر ضرورت ها و اقتضاهای عصر، انجام آن را کاملا توجیه میکند. به عنوان مثال، در مقر سازمان ناتو در بروکسل و جامعه اروپا علیرغم آنکه حدود ۱۲۰۰ مترجم ورزیده به کار اشتغال دارند، درحال حاضر از ترجمه ماشینی نیز استفاده می شود. دلیل این امر سرعت و هزینه است. میزان کاری که مترجمی ورزیده در خلال چندین روز انجام می دهد، توسط کامپیوتر در عرض چند دقیقه انجام می شود. حتی اگر کیفیت و دقت ترجمه ماشینی کمتر از حاصل کار مترجم باشد، باز هم از جهات گوناگون اهمیت و ارزش خاص آن چشمگیر است.

۳- بازیابی اطلاعات: اساس کار مراکز و پایگاههای اطلاع رسانی بر جستجو و بازیابی اطلاعات بنا شده است. زیر بخشی از زبانشناسی رایانه ای که به معناشناسی رایانه ای موسوم است تاکنون توانسته کیفیت این گونه سیستم ها را به نحو مطلوبی ارتقاء دهد. از معناشناسی منطقی نیز استفاده های فراوانی به عمل می آید.

به اختصار می توان گفت که اساس کار سیستمهای بازیابی اطلاعات اسناد موجود نیست، بلکه بازنمودهای معنایی است و ایجاد این بازنمودها کاری است در حیطه معناشناسی رایانه ای. در ایجاد زمینه زبانشناسان در ایجاد بازیابی غیر نظام مند، بازیابی نظام مند، ماشین های پرسش و پاسخ و ابهام زدایی از مفهوم واژه در متن نقش ایفا می نماید. دو مورد از مشکلات عمده موجود در سیستم های بازیابی

اطلاعات به ابهام و تشابه واژگانی مربوط است که رفع این مشکلات نیز در حیطه زبانشناسی رایانه ای است.

منطق توصیفی، شبکه های معنایی و صورت گرای قالب- محور مبنای کار سیستمهای اطلاع رسانی در حیطه معناست.

۴- بازشناسی گفتار و سیستم های گفتار به متن: بازشناسی گفتار بر

اساس مدلهای زبانشناختی بنا شده که از آن جمله می توان به مدل مارکو اشاره نمود. دستور حالت‌های محدود چامسکی نیز کاربرد فراوان دارد. در سیستم های بازشناسی گفتار شناخت امواج صوتی، چگونگی تعبیر و تفسیر آنها، رمزگذاری و رمز گشایی و چگونگی تغییر مشخصه های آوایی مسایلی عمده و اساسی می باشند که یک آواشناس یا واج شناس از عهده آن بر می آید. (آواشناسی و واج شناسی یکی از زیر شاخه های علم زبانشناسی است). تاکنون در زمینه تلفظ و املا الگوهای مختلفی ارائه گردیده است.

۵- برچسب زنی کلمات: این حیطه نیز از حیطه های مطرح در علم

اطلاع رسانی است و به یکی دیگر از زیر شاخه های علم زبانشناسی که نحو نام دارد مربوط است. شناخت مقوله های مختلف دستوری و چگونگی ایجاد ارتباط و تمایز بین دسته های مختلفی از کلمات به اطلاعات تخصصی در حیطه نحو نیازمند است.

۶- سیستم های متن به گفتار: حیطه رایج دیگری در علم اطلاع رسانی

است که از واج شناسی و آواشناسی رایانه ای استفاده های زیادی به عمل می آورد. شناخت اندام های صوتی، آوانویسی آواهای زبانی، استفاده از الگوها و قواعد

واجبی و واج شناسی در سیستم های متن به گفتار همگی نیازمند به فردی است که در حیطه واج شناسی به مطالعه و تحقیق پرداخته باشد. در ادامه برای رعایت اختصار کاربردهای عمده دیگر این حیطه فهرست وار ارایه می گردد.

۷- واژه سازی: جستجو و پیشنهاد معادلهای مناسب برای واژه های فاقد معادل فارسی (البته برای این کار لازم است با هماهنگی و یا همکاری فرهنگستان زبان و ادب فارسی اقدام گردد).

۸- نمودارهای N: از این نمودارها برای شمارش اعداد در متن و ... استفاده می شود و مطالب موجود تا حد زیادی به ریاضیات و علم رایانه نیز مربوط است.

۹- دستورهای بافت وابسته و مستقل از بافت: سیستم های بازیابی با استفاده از این گونه الگوها می توانند به جستجو یا تجزیه متن دست بزنند که از آن جمله می توان به تجزیه های از کل به جز و از جز به کل اشاره نمود.

۱۰- تجزیه و تحلیل معنایی

۱۱- گفتمان: که در آن در خصوص چگونگی تعبیر و تفسیر ضمایر و یافتن مرجع آنها و نکات مرتبط دیگر بحث می شود که در ماشینهای ترجمه و ... کاربرد فراوان دارد.

۱۲- تبدیل متن به گفتار

۱۳- تبدیل گفتار به متن

۱۴- خلاصه سازی خودکار

۱۵- موتورهای کاوش هوشمند

اهداف مهم گروه پژوهشی زبانشناسی رایانه ای بشرح زیر می باشد:

۱- مطالعه ساختار و واژگان های علمی و پژوهشی جهت دریافت توصیفگر های

مناسب

۲- مطالعه و بررسی علوم اطلاع رسانی و کتابداری و زبانشناسی و افزایش بهره

وری آنها در سطح کشور و منطقه

۳- مکانیزه کردن نمایه سازی و چکیده نویسی مدارک علمی کشور و منطقه

۴- مطالعه در ساختار زبان فارسی و ارتباط منطقی واژه ها و پراکنش آماری آنها

۵- مطالعه در ساختار ترجمه در زبانهای مختلف با استفاده از رایانه و منطق

زبانشناسی

۶- افزایش بهره وری اطلاع رسانی با استفاده از علم زبانشناسی در سطح کشور و

منطقه

۷- تعیین استراتژیهای جستجو با استفاده از اصلاحنامه ها و تزاروسها و

واژه نامه های کامپیوتری

۸- بررسی چگونگی تثبیت اصطلاحات فارسی و کاربرد صحیح آن زبان بعنوان

یک زبان علمی

۹- بررسی و رفع مشکلات جستجو و بازیابی اطلاعات رایانه ای با استفاده از

منطق و علم زبانشناسی

۱۰- بررسی ارتباط زبانشناسی و بازیابی اطلاعات با الگوهای منطقی زبانشناسی

۱۱- بررسی املای هوشمند از طریق ارتباط علم زبانشناسی و رایانه

- ۱۲- مطالعه ساختار و صرف و نحو زبانهای طبیعی با استفاده از علم منطق
زبانشناسی
- ۱۳- بررسی چگونگی ساختار (میانجی) واسط‌های زبانهای طبیعی برای
سیستمهای کامپیوتری
- ۱۴- بررسی استراتژی‌های تبدیل گفتار به نوشتار و بالعکس با استفاده از علم و
منطق زبانشناسی
- ۱۵- مطالعه خلاصه سازی خودکار در رایانه با استفاده از علم و منطق زبانشناسی
- ۱۶- بررسی چگونگی رو در رویی (ارتباط) زبان طبیعی و سیستمهای کامپیوتری
- ۱۷- برچسب زنی کلمات جهت بازشناسی رایانه‌ای، با استفاده از منطق زبانشناسی
- ۱۸- بررسی ساختار دستوره‌های بافت وابسته و مستقل از بافت جهت تعیین
الگوهایی برای بازیابی اطلاعات.

طرح‌های تحقیقاتی

پایان یافته در سال ۱۳۸۹

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

مجری: شاپور رضا برنجیان

تاریخ شروع: ۸۷/۱۰/۱

تاریخ پایان: ۸۹/۸/۳۰

چکیده:

تقریباً تمامی زبانهای موجود در دنیا در هر برهه از زمان بدلیل ماهیت خود زبان اقدام به واژه سازی می کنند، این عمل از دو طریق (اشتقاق و تصریف) صورت می گیرد، که اکثر آنها از بن های مضارع یا ماضی افعال ساخته می شوند. مانند: خواستار، گفتار، کردار، دانش ها و ... و همچنین صرف افعال فارسی (در تمام زمانها) نیاز به مشخص نمودن بن ماضی و مضارع افعال دارند، از این رو ساخت ریشه یا بن های ماضی و مضارع در افعال فارسی در ذخیره سازی و بازیابی اطلاعات، سیستمهای ماشینیهای ترجمه، تبدیل متن به گفتار، گفتار با متن و ... از اهمیت خاصی برخوردار است.

طرحهای در دست اجرا در سال ۱۳۸۹

ریشه یاب ماشینی افعال بسیط لازم

مجری: شاپور رضا برنجیان

تاریخ شروع: ۱۳۸۹/۹/۱۴

چکیده:

تحلیل خودکار متن و بازیابی رایانه ای اطلاعات موضوعی است که سالها پیش مطرح و موجب توجه زبانشناسی رایانه ای به ویرایش و پیرایش در مورد زبانها بویژه زبانهای انگلیسی، فرانسه و... گردید و همچنین موجب پیشرفت زیادی در ایجاد هماهنگی بین این زبانها با رایانه گردیده است که از آن جمله می توان به ساده کردن قواعد و کم کردن پیچیدگی آنها و یکسان کردن گفتار و نوشتار و گسترش استانداردهای آماده شده اشاره کرد.

فن هماهنگی بین زبانها و رایانه موجب تسهیل در بوجود آمدن ریشه یابها و ماشینهای ترجمه و... گردید. برای دستیابی به ریشه یابها و ماشینهای ترجمه ابتدا باید اجزاء مختلف تشکیل دهنده واژه ها با دقت نظر مورد مطالعه و بحث و بررسی قرار گیرند، از این رو طرح "ریشه یاب ماشینی افعال ساده و لازم در زبان فارسی در زمانهای گذشته، حال و آینده" جهت یاری رساندن به ساخت ماشینهای تحلیل خودکار متن و بازیابی اطلاعات در مدارک ذخیره شده به زبان فارسی ارائه می گردد.

بررسی چالشهای پردازش زبان طبیعی فارسی

مجری: دکتر حمید علیزاده

تاریخ شروع: ۱۳۸۸/۱۲/۲۷

چکیده:

پردازش زبان طبیعی، به فرایندها و پردازش هایی اطلاق می شود که در طی آن رایانه ها به گونه ای تعلیم داده می شوند که قادر به ادراک و تعبیر و تفسیر متون نوشته شده به زبان طبیعی می گردند. این پردازش ها منجر به ساخت ابزاری می گردد که در بازیابی بهینه اطلاعات بسیار راهگشا می باشد.

اگرچه تحقیقات زیادی در مورد فنون و ابزارهای مربوط به پردازش زبان طبیعی در سایر زبان ها انجام شده است، اما این مورد در زبان فارسی چندان با استقبال مواجه نشده و حجم پژوهش های مربوط در مقایسه با زبانهایی چون انگلیسی بسیار اندک است.

در این پژوهش ضمن بررسی حوزه های مربوط به پردازش زبان طبیعی، پژوهش های انجام شده معرفی و ضمن بررسی کاستی های موجود، چالش های هر حوزه معرفی می گردد، این چالشها می تواند راه را برای پژوهشهای آتی در حوزه پردازش زبان طبیعی هموار سازد.

سامانه انتقال اطلاعات از فایل XML نشریات فارسی
به سیستم جامع RICEST

مجری: مهندس محمدباقر دستغیب

تاریخ شروع: ۱۳۸۹/۱۱/۸

چکیده:

واژه نامه برگردان نام و نام خانوادگی نویسندگان خارجی به
فارسی با استفاده از تحلیل رخداد - محور

مجری: دکتر محمدرضا فلاحتی فومنی

تاریخ شروع: ۱۳۸۹/۸/۹

چکیده:

مقالات

نظری بر نواقص خط فارسی

نویسنده: شاپور رضا برنجیان

ارایه شده در فصلنامه علمی- پژوهشی اباخر سال پنجم بهار و تابستان شماره ۱۹

و ۲۰ اسفند ۱۳۸۹

چکیده:

پیدایش خط در هر جامعه ای موجب پیدایش تمدن و فرهنگ و توسعه علوم و فنون در آن جامعه بوده است خط عبارت است از علائمی که فکر و گفتار بشر را با استفاده از نشانه های دیداری ثبت و ضبط می کند. پس هر اندازه این علامت ها روشن تر، ساده تر، رسا تر، یکدست تر، و تعدادشان به تعداد اصوات زبان نزدیک تر باشد، سودمند تر است. در این مقاله ابتدا مختصری از تاریخچه خط در کشورمان ارائه شده و سپس اشکالات خط فارسی میانه و فارسی نو بطور کامل مورد بررسی قرار گرفته است و در نهایت چهار ویژگی اساسی یک خط جهت سادگی املاي آن ذکر گردیده است.

کلید واژه ها:

خط فارسی، اشکالات خط، تاریخچه خط، نواقص خط فارسی.

مقدمه ای بر Blade Server

نویسنده: مهندس محمدباقر دستغیب

ارائه شده در مجله وب شماره ۱۲۵ و ۱۳۸۹

چکیده:

سرورهای Blade سرورهایی هستند که از نظر فضا و مصرف انرژی بهینه سازی شده اند تا فضای کمتری را اشغال کنند و انرژی کمتری مصرف کنند. بطور کلی سرورهای Rack-mount فضا و کابل شبکه و برق کمتری مصرف می کنند. در کیس یا محفظه Blade چندین Blade سرور را می توان نصب کرد. معماری این سرورها بدین صورت می باشد که چندین سرور را درون یک محفظه قرار می دهد. این سرور در مراکز داده (data center) به منظور کاهش فضا و بهبود مدیریت سیستم مورد استفاده قرار می گیرد. هر دو نمونه ایستاده یا سوار شده بر روی rack توسط خود محفظه تأمین برق شده و هرکدام از blade ها پردازنده، حافظه و در صورت نیاز هارد دیسک مخصوص به خود دارند. علاوه بر این موارد تأمین نیروی برق می تواند اختیاری باشد.

بازیابی اطلاعات در مدارک XML

نویسنده: مهندس محمدباقر دستغیب

چاپ شده در : مجله وب، شماره ۱۲۴ و ۱۳۸۹

چکیده:

Inter-index consistency (IIC) in a Persian Context

Mohammad Reza Falahati Qadimi Fumani
The International Journal of Indexing
Volume 28, Number 1, March 2010

Abstract:

The author describes research he has carried out to assess the consistency (IIC) between pairs of indexers working on a collection of 111 Persian articles, using the Hooper, Rolling and cosine evaluation measures. By and large the IIC scores have been very low. He asks why this should be so, and to remedy the problem recommends better selection and training of indexers, better work conditions, and a collaborative approach to the task. He also points to the possibility that automatic indexing may offer a way forward.

The Persian Agrovoc in an indexing Context

Mohammad Reza Falahati Qadimi Fumani
The International Journal of Indexing
Volume 29, Number 1, March 2011

Abstract:

The author explores the Persian Agrovoc thesaurus from the point of view of the indexer, both human and automatic. It is found wanting, with problems both in the English version and in the Persian translation.

بررسی کارآمدی روش های موجود در بازیابی اطلاعات بین زبانی فارسی - انگلیسی با استفاده از واژه نامه: دو زبانه ماشین خوان

نویسنده: دکتر حمید علیزاده

چاپ شده در : فصلنامه علوم و فناوری اطلاعات ایران، دوره ۲۵، شماره ۱، صفحه

۷۰-۵۳، پائیز ۱۳۸۸

چکیده:

در این پژوهش میزان تأثیر انجام پردازش های زبان طبیعی بر روی ترجمه عبارت های جستجو با آزمون فرضیه های پژوهش مشخص گردید. فنون پردازش زبان طبیعی که برای پردازش عبارت های جستجو به کار گرفته شد شامل قطعه بندی متن، شناخت گونه های زبانشناختی، حذف سیاهه بازدارنده، تحلیل مورفولوژیک، و برچسب زنی انواع نقش دستوری بود. آزمون فرضیه اول نشان داد که استفاده از روش ترجمه اولین برابر نهاده در مقایسه با شیوه انتخاب همه برابر نهاده ها موجب کارآمدی بیشتر در بازیابی می گردد. آزمون فرضیه دوم نشان داد که اگرچه تحلیل مورفولوژیک واژه هایی که به وسیله واژه نامه ترجمه نشدند باعث افزایش ضریب دقت بازیافت می گردد، اما تفاوت معناداری با عدم انجام این تحلیل ایجاد نمی نماید. بررسی فرضیه سوم نیز نشان داد که ترجمه عبارتی در مقایسه با ترجمه واژه به واژه باعث کارآمدی پیش تر می گردد. یافته های دیگر این پژوهش نیز نشان داد که دگرنویسی واژه های فارسی ترجمه ناپذیر با حروف انگلیسی و قراردادن آن ها در عبارت جستجوی نهایی در مقایسه یا حذف آن ها از عبارت های جستجو، می تواند منجر به افزایش کارآمدی گردد.

کلید واژه ها: بازیابی اطلاعات بین زبانی، پردازش زبان طبیعی، واژه نامه دوزبانه ماشین خوان، ارزیابی بازیابی اطلاعات.

سخنرانی‌ها

Familiarity with RICEST

سخنران: دکتر محمدرضا فلاحتی قدیمی فومنی

مکان سخنرانی: تهران سمینار ISC

زمان: اکتبر و نوامبر 2010 (۱۰-۲ آبان ماه ۱۳۸۹)

FaNexer: Persian Automatic Indexer

سخنران: دکتر محمدرضا فلاحتی قدیمی فومنی

مکان سخنرانی: شیراز، مرکز منطقه ای اطلاع رسانی علوم و فناوری -

همایش هفته پژوهش

زمان: دسامبر 2010 (۹-۴ دی ماه ۱۳۸۹)

آشنایی با ISC-ESI

سخنران: دکتر حمید علیزاده

مکان سخنرانی: تهران، نمایشگاه مطبوعات تهران - همایش ISC

زمان: اکتبر ۲۰۱۰ (مهرماه ۱۳۸۹)

چکیده: ESI از جمله زیر سیستم های پایگاه استنادی علوم جهان اسلام است که به بررسی برترین موالفین، نشریات، دانشگاهها و موسسات و همچنین مقالات داغ و جبهه های تحقیق می پردازد. در این سخنرانی با مبانی علمی این حوزه و امکانات این پایگاه آشنا شده و با ارائه توضیحات براساس نمایش این پایگاه، شیوه بهره گیری از آن در پژوهش های علم سنجی ارائه می گردد.