

چکیده

وب فعلی برای استفاده انسان ایجاد شده تا قادر به مشاهده اطلاعات مورد نیاز خود و فهم آنها باشد. وب معنایی در کنار این مهم، تعامل ماشین با محتویات وب و درک خودکار مفاهیم آن را مورد توجه قرار داده است. در این وب، محتویات بایستی بگونه ای قابل فهم برای ماشین نمایش داده شوند. عمده کارها در این حوزه به دو دسته استخراج مفاهیم و استخراج روابط تقسیم می‌شوند.

بسیاری از کاربردها در استخراج اطلاعات، اکتساب دانش، فهم زبان طبیعی و بازیابی اطلاعات نیازمند درک در سطح مفهوم، معنا، رابطه و روابط معنایی می‌باشد. هر کدام از موجودیتهای ذکر شده دارای ویژگیها و تعاریف مشخص می‌باشند. رابطه معنایی به روابط بین مفاهیم موجود در ذهن که به آن مفاهیم ادراکی یا به روابط بین کلمات که روابط لغوی نامیده می‌شوند، اشاره دارد.

روابط معنایی در حوزه متن می‌توانند در سطوح پایین و بین واژگان و یا در سطوح بالاتر بین عبارات، جمله ها، پاراگرافها و حتی بالاتر از بخشهای یک متن مثلا بین دو سند یا مجموعه ای از اسناد رخ دهد. سطوح ذکر شده دارای یک محدوده تعریف ساختاری می‌باشند؛ اما سطح دیگری در حوزه زبان شناسی که علی‌رغم اهمیت بالا، کمتر در حوزه زبان شناسی رایانه ای به آن پرداخته شده است سطح گفتمان نام دارد؛ که محدوده تعریف آن نه ساختار بلکه معنا است. در این سطح، مقیاس نه یک جمله یا پاراگراف یا سند بلکه واحد معنایی است؛ و ممکن است بین چند جمله، چند پاراگراف و یا تمام متن باشد.

نکته قابل توجه وقوع این روابط معنایی در سطح گفتمان، بصورت صریح و یا ضمنی در متن می‌باشد و اهمیت آن به دلیل انتشار روزافزون اطلاعات در سطح وب است که غالبا در قالب متن هستند و اکتساب دانش و استخراج اطلاعات مورد نیاز از بین این حجم انبوه داده مستلزم درک مفاهیم و روابط معنایی حاکم بین آنها بویژه در حوزه های زبان شناسی، روانشناسی، علوم شناختی و در این رساله زبان شناسی محاسباتی است.

این رساله روشی جدید محاسباتی به منظور استخراج و نمایش روابط معنایی صریح و ضمنی در سطح گفتمان از متن ارائه داده است که مکمل و توسعه دهنده روشهای پیشین است. بهره گیری از روشهای پردازش زبان طبیعی در کنار محاسبات آماری و روش تجزیه مقدار تکینه و روش استخراج و شناسایی الگو بدون ناظر از مشخصه های روش پیشنهادی است. هدف کمک به درک و فهم دقیق و عمیق تر از متن و پیاده سازی در کاربردهای بسیار در حوزه های مرتبط با زبان شناسی، روانشناسی، علوم شناختی و بویژه زبان شناسی محاسباتی است. تحقیقات موجود نشان می دهد که روشهای ارائه شده هوش مصنوعی و یادگیری ماشین در این حوزه نیازمند به سطح بالایی از مهندسی دانش یا پایگاه داده های بزرگ برای یادگیری و آزمایش هستند. روش ارائه شده ضمن نوآوری در حوزه پیشنهادی، محدودیت های و مشکلات ماقبل را حل نموده است. در این پایان نامه توسعه، پیاده سازی و ارزیابی روش برای محاسبه کارآیی، دقت و صحت انجام پذیرفته است.

روش محاسباتی پیشنهادی از شیوه جبر خطی تجزیه مقدار تکینه در کنار انجام برخی پردازشها و پیش پردازشها در حوزه زبان طبیعی از جمله برچسب گذاری نقش معنایی، برچسب زنی نحوی و تکنیکهای آماری پردازش زبان طبیعی و روشهای محاسبه پیوستگی و همبستگی نیز بهره گرفته است.

کاربردهایی در ایجاد هستان شناسی و تکامل و توسعه آن، استخراج دانش، سیستم های پرسش و پاسخ، خلاصه ساز متن و بازیابی اطلاعات مانند موتورهای جستجو از آن جمله می باشد.

ارزیابی روش پیشنهادی با دو رویکرد صورت پذیرفته است: (۱) اعمال روش بر روی داده های استاندارد و (۲) پیاده سازی روش پیشنهادی در موارد حوزه کاربرد. ارزیابیهای حاصل نشانگر عملکرد قابل توجه روش پیشنهادی دارد.

کلمات کلیدی: سطح گفتمان، روابط معنایی، تجزیه مقدار تکینه، اکتساب دانش، پردازش زبان طبیعی،

زبان شناسی محاسباتی، محاسبات آماری، حوزه معنایی