

طراحی و پیاده سازی یک سامانه ترجمه فارسی به انگلیسی

فائزه میرزائی^۱، محسن بیگلری^۲، احمد یوسفان^۳ و عماد بیات^۴

دانشگاه کاشان؛ دانشکده مهندسی، گروه مهندسی کامپیوتر

fmirzaei@grad.kashanu.ac.ir^۱

biglari@grad.kashanu.ac.ir^۲

yoosofan@kashanu.ac.ir^۳

emadbayat85@yahoo.com^۴

چکیده

ترجمه‌ی خودکار ماشینی از شاخه‌های فعال پردازش زبان طبیعی در سال‌های اخیر بوده است. هر سیستم پردازش زبان طبیعی نیاز به یک بخش پیش‌پردازش برای تبدیل ورودی به شکل مناسب دارد. این پیش‌پردازش می‌تواند شامل جداسازی عبارت ورودی به جملات، جملات به کلمات و اصلاح املا کلمات در صورت نیاز باشد.

در این مقاله یک سیستم ترجمه‌ی فارسی به انگلیسی با استفاده از روش مبتنی بر قوانین ارائه شده است که بخش پیش‌پردازش را نیز شامل می‌شود. در این سیستم بخش مهمی از پردازش‌ها به زمان طراحی سیستم منتقل شده است؛ و به موجب این انتقال، سرعت و دقت سیستم در تجزیه‌ی جملات افزایش چشمگیری پیدا کرده است. برای واژگان فارسی لغت‌نامه دهخدا به کار گرفته شده و برای ترجمه‌ی انگلیسی یک فرهنگ واژگان دوزبانه با نزدیک به ۴۵۰۰۰ لغت به کار گرفته شده است. در زمان اجرای سیستم جاری، تنها نیاز به تشخیص مواردی چون نشانه‌های جمع، زمان افعال و ضمائر متصل مفعولی می‌باشد. سیستم ارائه شده به صورت کامل پیاده‌سازی و ارزیابی شده است و نتایج بدست آمده، نشان داده است که چنین روشی کاملاً کارآمد است.

کلمات کلیدی

ترجمه‌ی ماشینی، ترجمه‌ی فارسی به انگلیسی، لغت‌نامه دهخدا، پردازش زبان طبیعی

روش‌های مبتنی بر قوانین: این روش‌ها از مجموعه‌ای از قوانین برای ترجمه بهره می‌برند که براساس دانش زبان‌شناسی انسان نوشته شده است.

روش‌های آماری: این روش‌ها، براساس مثال‌هایی از ترجمه‌ی انجام شده توسط انسان سعی بر یادگیری ترجمه دارند.

۲- مشکلات ترجمه فارسی به انگلیسی

گام اول در ترجمه، تجزیه‌ی جملات زبان مبدا است؛ از دشواری‌های تجزیه‌ی جملات زبان فارسی، می‌توان به موارد زیر اشاره کرد:

- بعضی از اسم‌ها دارای چند معنی متفاوت هستند، مانند شیر.
- بطور معمول، حروف صدادار کوچک در نوشتار فارسی ظاهر نمی‌شوند که موجب ابهام در کلمات متشابه می‌شود. برای مثال، کلمه‌ی «گل» را می‌توان به دو صورت «گِل» و «گُل» تفسیر کرد.

۱- مقدمه

ترجمه‌ی یک متن از یک زبان طبیعی به زبانی دیگر، ترجمه‌ی ماشینی نامیده می‌شود. در ترجمه‌ی خودکار ماشینی، عملیات ترجمه توسط کامپیوتر صورت می‌گیرد. این شاخه از پردازش زبان طبیعی در ۵۰ سال اخیر، موضوع پژوهش بسیاری از محققان بوده و توجه زیادی را به خود جلب کرده است. روش‌های ارائه شده را می‌توان به صورت کلی به چند بخش تقسیم کرد:

روش‌های مبتنی بر فرهنگ واژگان: در این روش‌ها، منبع اصلی ترجمه یک فرهنگ واژگان دوزبانه است که بسیار تعیین کننده بوده و قدرت ترجمه نیز تا حدود زیادی به آن بستگی دارد.

روش‌های مبتنی بر دانش یا مثال: در این روش‌ها، با استفاده از دانش و نوشتجات دوزبانه (مثال‌ها)، تک‌تک موارد از زبان مبدا به زبان مقصد نگاشت می‌شوند.