



## گزارش نهایی طرح پژوهشی

عنوان

بررسی ساختار علمی و تکامل حوزه ترجمه ماشینی در سال‌های ۲۰۰۶ تا ۲۰۱۵ بر

اساس داده‌های پایگاه اسکوپوس

مجری

دکتر حمید علیزاده

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

## فهرست مندرجات

۶	فصل یک
	۱-۱ مقدمه ۸
۱۱	۱-۲ بیان مساله
۱۲	۱-۳ ضرورت انجام پژوهش و کاربرد نتایج
۱۳	۱-۴ اهداف پژوهش
۱۳	۱-۵ سوالات پژوهش
	۱-۶ تعاریف مفهومی و عملیاتی
۱۵	فصل دوم
۱۵	مبانی نظری و مرور پیشینه‌های پژوهش
۱۶	۱-۲ مقدمه
۱۶	۲-۲ مبانی نظری
۱۶	۱-۲-۲ زبان علم و ترجمه ماشینی
۲۱	۲-۲-۲ علم سنجی و تحلیل هم‌واژه
۲۳	۲-۲-۳ سنجش و ارزیابی و ترجمه ماشینی
۲۴	۲،۳- مرور پیشینه‌های بین‌المللی
۲۷	۲،۴- مرور پیشینه‌های پژوهش‌های ملی
۲۷	۲،۵- استنتاج از پیشینه پژوهش
۲۹	فصل سوم
۲۹	روش‌شناسی پژوهش
۳۱	۳-۱ مقدمه
۳۱	۳-۲ روش پژوهش
۳۱	۳-۳ نوع پژوهش
۳۱	۳-۴ جامعه پژوهش
۳۲	۳-۵ روش گردآوری داده‌ها

فصل چهارم ..... ۳۴

تجزیه و تحلیل داده‌ها ..... ۳۴

۴-۱. مقدمه ..... ۳۵

۴-۲. سوال اول پژوهش: مولفین بر تولید حوزه ترجمه ماشینی چه کسانی هستند؟ ..... ۳۵

۴-۳. سوال دوم پژوهش: دانشگاه‌ها و کشورهای برتر در انتشارات حوزه ترجمه ماشینی در سال‌های ۲۰۰۵ تا

۲۰۱۶ کدامند؟ ..... ۳۶

۴-۵. سوال سوم پژوهش: توزیع موضوعی انتشارات ترجمه ماشینی چگونه است؟ ..... ۳۸

۴-۶. سوال چهارم پژوهش: نقشه‌های علمی در خصوص ساختار و تکامل رشته ترجمه ماشینی چه اطلاعاتی

ارائه می‌کنند؟ ..... ۳۹

فصل پنجم ..... ۴۳

نتیجه‌گیری و پیشنهادها ..... ۴۳

۵-۱. مقدمه ..... ۴۵

۵-۲. بحث و نتیجه‌گیری ..... ۴۵

۵-۳. نتیجه‌گیری کلی ..... ۴۸

۵-۴. پیشنهادهایی برای پژوهش‌های آینده ..... ۴۹

۵-۵. پیشنهادهای کاربردی و برگرفته از نتایج پژوهش حاضر ..... ۵۰

فهرست منابع ..... ۵۱

## چکیده

**زمینه و هدف:** این پژوهش باهدف شناسایی گرایش‌های پژوهشی اصلی حوزه ترجمه ماشینی انجام شد. ترجمه ماشینی حوزه‌ای میان‌رشته‌ای است که هدف آن ترجمه متون علمی در یک زبان با متن مشابه در زبانی دیگر است. برای نیل به هدف پژوهش از دیداری‌سازی حوزه‌های دانش، برای نمایش شبکه واژگان عناوین و چکیده‌های مقالات ترجمه ماشینی استفاده شد.

**روش‌شناسی:** این پژوهش کاربردی است و در آن از روش علم‌سنجی و دیداری‌سازی استفاده شده است. در این پژوهش انتشارات پژوهشی این حوزه در سال‌های ۲۰۰۶ تا ۲۰۱۵ به روش متن‌کاوی مورد تحلیل قرار گرفت. پیکره پژوهش از انتشارات علمی این حوزه در پایگاه اسکوپوس استخراج گردید. متن‌کاوی با استفاده از روش هم‌رخدادی واژه‌ها انجام شد. تحلیل هم‌رخدادی واژه‌ها یکی از روش‌های تحلیل محتوا است که در این پژوهش روی پیکره انتشارات حوزه ترجمه ماشینی اجرا شد. این روش بینشی کامل نسبت به ساختار علمی و گرایش‌های پژوهشی این حوزه ارائه نمود.

**یافته‌ها:** در نهایت نتایج تحلیل متن کاوی، ساختار علمی این حوزه را بر اساس خوشه‌های موضوعی آن شناسایی کرد. یافته‌های پژوهش شامل شناسایی چهار خوشه موضوعی است که گرایش‌های اصلی تحقیقاتی در این حوزه بشمار می‌آیند. در این میان دو خوشه برتر تحت عنوان ترجمه ماشینی آماری و ترجمه خودکار و ارتباطات کاربران به‌عنوان گرایش‌های اصلی پژوهشی حوزه ترجمه ماشینی معرفی شد.

**نتیجه‌گیری:** علی‌رغم اهمیت حوزه ترجمه ماشینی، تحقیقات علم‌سنجی به شیوه مؤثر در این حوزه انجام نشده است. لازم است با انجام این تحقیقات و اجرای آنها با سایر رویکردهای دیداری‌سازی، اطلاعات بیشتری در خصوص ساختار و تکامل این رشته در اختیار محققان قرار گیرد.

**کلیدواژه‌ها:** ترجمه ماشینی، متن کاوی، ساختار علم، تحلیل هم‌واژه

## فصل یک



## ۱-۱ مقدمه

هدف ترجمه ماشینی، ترجمه زبان‌های انسانی به شکلی خودکار است که کاربر را قادر سازد از این زبان‌ها و منابع نوشتاری آنها بهره‌بردارد. فرآیند ترجمه ماشینی یک نیاز ضروری برای رسیدگی به موانع زبانی و توجه به نابرابری‌های اطلاعاتی در میان مردمانی است که به حدود ۶۵۰۰ زبان دنیا صحبت می‌کنند. استفاده از زبان انگلیسی به عنوان زبان علمی تنها ۴۰۰ سال پیش در کنار امپراتوری استعماری رو به رشد بریتانیای کبیر آغاز شد. پس از جنگ‌های جهانی، با افزایش نفوذ نظامی، اقتصادی و فناوری ایالات متحده به گسترش خود ادامه داد (پییم، ۲۰۱۴). از آن زمان، تسلط انگلیسی در سراسر جهان گسترش یافته است، البته در کنار آن تاثیر فناوری اطلاعات و ارتباطات که ابزار آن عموماً بر اساس زبان انگلیسی طراحی شده است نیز موثر بوده است. امروزه، ۹۸٪ از انتشارات علمی مورد بررسی به زبان انگلیسی است (بلی و جردن، ۲۰۰۳) و انگلیسی زبان رسمی اکثر رویدادهای علمی و مجلات بین‌المللی و دانشگاهی نمایه شده است.

داشتن یک زبان مشترک با تسهیل ارتباطات علمی بین‌المللی و ایجاد یک مخزن تک‌زبان برای انتشارات و داده‌ها به علم کمک می‌کند (هاچینز، ۲۰۰۵). زبان انگلیسی اکنون زبان غالب ارتباطات علمی است. استفاده از منابع این زبان و منابع سایر زبان‌ها نیازمند نظام‌های ترجمه ماشینی قدرتمند است. برای کامل و معنی‌دار بودن ترجمه، درک عمیق ویژگی‌های نحوی، معنایی و صرفی که مشخصه یک جفت زبان هستند، یک پیش‌نیاز مهم را تشکیل می‌دهد. از جمله نظام‌های خودکاری که در این زمینه به یاری کاربر شتافته تا بهره‌مناسبی از منابع سایر زبان‌ها ببرد، نظام ترجمه ماشینی است. ترجمه ماشینی از طریق یک برنامه نرم‌افزاری بدون کمک انسانی انجام می‌شود. این حوزه به عنوان زیرشاخه‌ای از زبان‌شناسی محاسباتی و مهندسی زبان که ایده‌ها و تکنیک‌های خود را از زبان‌شناسی، علوم کامپیوتر، هوش مصنوعی، تئوری ترجمه و آمار می‌گیرد، تکامل یافته است.

از سوی دیگر ترجمه ماشینی را می‌توان زیرشاخه‌ای از حوزه پردازش زبان طبیعی دانست (هاچینز، ۲۰۰۵). علوم دیگر از جمله فناوری اطلاعات و زبان‌شناسی نیز در تحقیقات آن مشارکت دارند.



مطالعات علم‌سنجی در حوزه ترجمه ماشینی از قدمت زیادی برخوردار نیست. یکی از دلایل این امر نوپا بودن حوزه ترجمه ماشینی در مقایسه با سایر حوزه‌های علمی است. اما رویکردهای نوینی که برای مطالعه نشریات دانشگاهی و مطالعات ترجمه ماشینی از دید علم‌سنجی در دهه اخیر انجام مطرح شده است، چشم اندازهای جدیدی را در این عرصه گشوده است. در حقیقت، با افزایش بهره‌گیری از امکاناتی که فناوری اطلاعات و ارتباطات در اختیار محققین و عامه مردم قرار داده است، فرایندهای تولید، ذخیره‌سازی و اشاعه اطلاعات دچار تغییرات شگرفی شده است. امروزه پایگاه‌های اطلاعاتی عظیم، میلیون‌ها رکورد اطلاعاتی در اختیار دارند. علاوه بر این شبکه جهانی وب سرشار از منابع متنی است که به جرات می‌توان گفت کمتر موضوعی را می‌توان یافت که جستجوی اطلاعات در وب، حجم عظیمی از منابع را در آن خصوص بازیابی ننماید. این حجم عظیم منابع متنی باعث شده است که تحلیل و پردازش متون به صورت دستی در محیط اطلاعاتی امروز کاری صعب و طاقت فرسا تلقی شود. نتایج بازیابی اطلاعات به این صورت نیز دیگر چندان قابل اعتماد و تعمیم نیست (نورود، ابلکس، ۲۰۰۹). استفاده از سیستم‌های خودکار تحلیل متن می‌تواند جوابگوی این نیاز بوده و به نحو شگفت‌انگیزی در وقت و انرژی کاربر صرفه جویی نماید.

تعداد فزاینده‌ای از انتشارات حوزه ترجمه ماشینی با هدف طبقه‌بندی جنبه‌های مختلف و توضیح توسعه تاریخی و وضعیت فعلی این حوزه انجام شده است (بیکر و مایر، ۲۰۱۱). علاوه بر موارد ذکر شده، امروزه پژوهش در حوزه ترجمه ماشینی در کشورهای خاص (بلی و جردن، ۲۰۰۳) و یا موضوعات خاص (گاسپاری و دیگران، ۲۰۱۵) از قبیل زبان‌شناسی پیکره‌ای (بیکر و مایر، ۲۰۱۱) نیز انجام شده است. تا امروز اغلب مطالعات علم‌سنجی در حوزه ترجمه ماشینی بر اساس پایگاه‌های اطلاعاتی کتابشناختی، مطالعه روابط استنادی، تاثیر کارآمدی ترجمه ماشینی گوگل، مطالعه شبکه‌های هم‌نویسندگی و مواردی از این دست بوده است (ووس و ژایو، ۲۰۰۵). همانگونه که حوزه‌های درگیر در توسعه دانش ترجمه ماشینی دارای تنوع می‌باشد، جامعه پژوهشی بزرگی نیز در سراسر دنیا به پژوهش در این حوزه و انتشار یافته‌های آن از طریق مجلات علمی و یا ارائه در کنفرانس‌های بین‌المللی مشغولند. برای

درک بهتر و ایجاد معیارهایی برای پژوهش، نیازمند انجام مطالعه علم سنجی و دیداری سازی نتایج هستیم. نقشه‌های علمی در این میان جایگاه ویژه ای دارند.

نقشه‌ها ابزاری قدرتمند برای شناسایی و توصیف تعامل بین حوزه‌های مختلف هستند. با استفاده از آنها ساختار حوزه‌های پژوهشی و تکامل رشته‌های علمی قابل مطالعه می‌گردد. با استفاده از نقشه‌های کتابسنجی، ارتباط بین مفاهیم اصلی این حوزه به صورت بصری نشان داده می‌شود. این شیوه همچنین گرایش‌های موجود در حوزه را هم نشان می‌دهد. ضمن آنکه موضوع‌های اصلی پژوهش در این رشته که اغلب پژوهشگران به آن توجه دارند مشخص می‌شود.

از شاخه‌های اصلی علم سنجی می‌توان به دیداری سازی اشاره کرد. دیداری سازی ابزاری حیاتی برای درک و به اشتراک گذاری بینش در مورد داده‌ها است. ارائه مستقیم و ملموس داده‌ها به محققین، می‌تواند فضایی را برای بررسی باز کند و به خوانندگان اجازه دهد تا فرضیه‌ها و سؤالاتی را که با آن مواجهند را بصورت دیداری بررسی و پاسخ گویند (جاکوبی و دیگران، ۲۰۰۸). با توجه به توانایی ابزارهای دیداری سازی برای ترکیب مقادیر زیادی از داده‌ها و تحلیل و نمایش آنها با گرافیک‌های موثر، دیداری سازی به یکی از مهم‌ترین مؤلفه‌های تحقیق در رشته‌های مختلف تبدیل شده است (هولزینگر و دیگران، ۲۰۰۸). تحلیل هم واژه که یکی از روش‌های دیداری سازی است توسط کالون و همکاران پیشنهاد شده است (ریپ و کورتیال، ۱۹۹۴). این یک روش تحلیل محتوا است که علم سنجی و متن کاوی را با هم ترکیب می‌کند در واقع یک فناوری برای آشکار کردن معنای عمیق اسناد (کالون و دیگران، ۱۹۸۳). پنج مرحله ضروری برای دیداری سازی هم واژه‌ها وجود دارد: جمع آوری متون، استخراج داده‌ها، پردازش داده‌ها، تجزیه و تحلیل داده‌ها و دیداری سازی نتیجه (یانگ، وو و کوی، ۲۰۱۲). با بهره‌گیری از شاخص‌های علم سنجی و اجرای تکنیک نگاشت و دیداری سازی پیکره انتشارات ترجمه ماشینی، ساختار علمی حوزه، شامل تاثیرگذارترین مولفین، دانشگاه‌ها و کشورهایی که در این حوزه فعالیت دارند مشخص می‌شود. نشریات برتر و کنفرانس‌هایی که در این حوزه معتبرتر هستند و پراستنادترین مقالات که نشان دهنده حوزه‌های برتر موضوعی ترجمه ماشینی است نیز مشخص می‌شود.

بدست آوردن بینش عمیق و دید کلی نسبت به ساختار علمی حوزه ترجمه ماشینی از طریق مصور سازی شبکه هم واژه و با استفاده از واژه‌هایی که بیشترین تناوب را در انتشارات زبانشناسی رایانه ای دارند، قابل نمایش و تحلیل و بررسی است.

## ۱-۲ بیان مساله

پژوهش در حوزه ترجمه ماشینی با شتاب و وسعت مثال زدنی به پیش می‌رود. جنبه‌های اقتصادی ترجمه ماشینی و نیاز به معیارهای جدید بررسی کارامدی، از حوزه‌های پژوهشی اصلی است که به موارد پیشین اضافه شده است. بنابراین نیازی ضروری برای بررسی وضعیت تولید علم و پژوهش در این حوزه و پیش بینی سمت و سوی تحقیقات آینده در آن احساس می‌شود.

حجم زیاد پژوهش‌ها و ابزار ترجمه ماشینی، نیاز به معیارهای ارزیابی در این حوزه را ضروری ساخته است (پیم، ۲۰۱۴). از جمله حوزه‌های پژوهشی مهم ترجمه ماشینی می‌توان به بررسی کارامدی این نظام‌ها در اجرای دقیق ترجمه متون اشاره نمود. بدیهی است که علاوه بر سرعت، کیفیت ترجمه ماشینی از نکات حائز اهمیت بشمار می‌رود. این افزایش کیفیت در طول زمان به مدد ارزیابی نظام‌های موجود و مشخص شدن نقاط ضعف آنها حاصل شده است. در جهت آشنایی با وضعیت کنونی تحقیق در این حوزه و شناسایی نقاطی که باید بر روی آن تمرکز داشت و به تحقیقات جدید دست زد، باید دید جامعی نسبت به گذشته، حال و پیش بینی آینده این حوزه داشت. چنین چشم اندازی توسط بررسی علم سنجی حوزه ترجمه ماشینی حاصل می‌شود. هم اکنون تصویر جامع و یکپارچه ای از روند انجام پژوهش در حوزه ترجمه ماشینی در سطح بین المللی وجود ندارد. حوزه‌های پژوهشی برتر که نیاز به تحقیق و بررسی دارد چندان مشخص نیست. این نقایص انجام این پژوهش را ضروری می‌سازد. با

شناسایی زیر شاخه‌های فرعی این موضوع، نویسندگان برتر، کشورهای پیشرو و مقالات تاثیرگذار در این حوزه می‌توان به شناسایی همکاران پژوهشی بالقوه پرداخت و با همکاری در انجام تحقیقات مشترک به ویژه در زمینه تولید و توسعه نظام‌های ترجمه ماشینی فارسی و سایر زبان‌ها اقدام کرد. علم سنجی به دنبال «تحلیل، کمی کردن و اندازه‌گیری پدیده‌های ارتباطی برای نمایش رسمی دقیق رفتار آنها و توضیح، ارزیابی و یا شناسایی ساختار آنها است (گارفیلد، ۲۰۰۱).

تکامل حوزه ترجمه ماشینی در طول زمان که نشان دهنده شیوه توسعه این رشته در طول زمان است و وارد و صادر شدن ایده‌ها از این رشته به رشته‌های دیگر نیز در این تحلیل مشخص می‌گردد. در نهایت نحوه و میزان مشارکت پژوهشگران در این حوزه نیز مساله دیگری است که در این پژوهش بررسی می‌شود. بدست آوردن بینش عمیق و دید کلی نسبت به ساختار علمی حوزه ترجمه ماشینی از طریق مصور سازی شبکه هم‌واژه و با استفاده از واژه‌هایی که بیشترین تناوب را در انتشارات زبانشناسی رایانه ای دارند، قابل نمایش و تحلیل و بررسی است. پژوهشگران این حوزه نیز با آگاهی از یافته‌های این تحقیق می‌توانند به گونه ای هوشمندانه سیر پژوهش ترجمه ماشینی در جهان را دنبال کرده و از تجربیات انجام شده در زبان‌های دیگر برای مقاصد بومی و ترجمه خودکار در زبان فارسی استفاده کنند.

تکامل حوزه ترجمه ماشینی در طول زمان که نشان دهنده شیوه توسعه این رشته در طول زمان است و وارد و صادر شدن ایده‌ها از این رشته به رشته‌های دیگر نیز در این تحلیل مشخص می‌گردد. در نهایت نحوه و میزان مشارکت پژوهشگران در این حوزه نیز مساله دیگری است که در این پژوهش بررسی می‌شود. بنابراین مساله اصلی این پژوهش شناسایی ساختار و تکامل این حوزه و حوزه‌های موضوعی اصلی و چالش‌های تحقیقاتی آن است.

## ۱-۳ ضرورت انجام پژوهش و کاربرد نتایج

هم اکنون تصویر جامع و یکپارچه ای از روند انجام پژوهش در حوزه ترجمه ماشینی در سطح بین المللی وجود ندارد. حوزه‌های پژوهشی برتر که نیاز به تحقیق و بررسی دارد چندان مشخص نیست. این نقایص انجام این پژوهش را ضروری می‌سازد. با شناسایی زیر شاخه‌های فرعی این موضوع، نویسندگان برتر، کشورهای پیشرو و مقالات تاثیرگذار در این حوزه می‌توان به شناسایی همکاران پژوهشی بالقوه پرداخت و با همکاری در انجام تحقیقات مشترک به ویژه در زمینه تولید و توسعه نظام‌های ترجمه ماشینی فارسی و سایر زبان‌ها اقدام کرد.

#### ۴-۱ اهداف پژوهش

این پژوهش دارای هدف عمده ای است که عبارت است از بررسی ساختار علمی و تکامل حوزه ترجمه ماشینی در سال‌های ۲۰۰۶ تا ۲۰۱۵. در کنار آن اهداف فرعی زیر نیز جستجو می‌گردد:

- آشنایی با مولفین پرتولید این حوزه
- شناسایی دانشگاه‌ها و کشورهای برتر در انتشارات این حوزه
- بررسی نقشه‌های علمی در خصوص ساختار و تکامل رشته ترجمه ماشینی

#### ۵-۱ سوالات پژوهش

در این پژوهش سوالات زیر مورد بررسی قرار می‌گیرد:

- ۱- مولفین پرتولید حوزه ترجمه ماشینی چه کسانی هستند؟
- ۲- دانشگاه‌ها و کشورهای برتر در انتشارات این حوزه کدامند؟
- ۳- توزیع موضوعی انتشارات ترجمه ماشینی چگونه است؟
- ۴- نقشه‌های علمی در خصوص ساختار و تکامل رشته ترجمه ماشینی چه اطلاعاتی ارائه می‌کنند؟



## فصل دوم

### مبانی نظری و مرور پیشینه‌های پژوهش

## ۱-۲. مقدمه

در این فصل ابتدا مبانی نظری با زبان علم و ترجمه ماشینی آغاز می شود. آنگاه علم سنجی و تحلیل هم واژه مطرح می شود. در نهایت با سنجش، ارزیابی و ترجمه ماشینی، بخش مبانی نظری به پایان می رسد. پس از آن مروری بر پیشینه پژوهش در سطوح ملی و بین المللی مطرح می شود. این بخش در نهایت با استنتاج از پیشینه پژوهش به پایان می رسد.

## ۲-۲ مبانی نظری

### ۱-۲-۲ زبان علم و ترجمه ماشینی

حفظ یک زبان علمی مشترک برای انتشار و به رسمیت شناختن تحقیقات انجام شده توسط دانشمندی که زبان اصلی آنها به طور گسترده صحبت نمی شود و همچنین برای تسهیل ارتباط بین این دانشمندان و جامعه علمی گسترده تر مفید است. داشتن یک زبان علمی مشترک همچنین تحرک بین المللی را تسهیل می کند و تعداد زبان های اضافی مورد نیاز برای همکاری بین المللی را محدود می کند. با این حال، علیرغم مزایای یک زبان مشترک، حفظ یک زبان علمی واحد و جهانی موانعی ایجاد می کند که اکثر محققان در جهان را ملزم می کند تا قبل از تعامل با جامعه دانشگاهی جهانی، در یک زبان دیگر مهارت پیدا کنند. یونسکو از طریق توصیه خود در مورد علوم باز، از مؤسسات علمی خواسته است تا برنامه های تحقیقات جهانی، چند زبانه و بین رشته ای را به منظور دسترسی عادلانه تر به دانش و کسب و کارهای علمی تقویت کنند.

از آنجا که دانش آکادمیک بیشتر به زبان انگلیسی منتقل می شود، دانشمندان و سایر اعضای جامعه، اغلب دانش تولید شده در زبان های دیگر را نادیده می گیرند، برای مثال تنها با استفاده از کلمات کلیدی انگلیسی جستجو انجام می دهند (هاچینز، ۲۰۰۵). نادیده گرفتن مطالعات غیرانگلیسی در موتورهای جستجو، می تواند منجر به شکاف های بزرگ در پایگاه های داده جهانی شود که بر سیاست، مدیریت و تصمیم گیری تأثیر می گذارد (آمانو و همکاران، ۲۰۱۶). مهارت انگلیسی همچنین بر افرادی که در علم در مقیاس جهانی مشارکت می کنند تأثیر می گذارد و برای علم مضر است، زیرا



تنوع دیدگاه‌ها ساخت دانش علمی قوی و نوآورانه را تقویت می‌کند. تسلط به زبان انگلیسی اغلب یک شرط لازم برای پیشرفت حرفه‌ای است، مانند انتشار در مجلات با تاثیر بالا، دریافت کمک‌های مالی بین‌المللی، یا شرکت در کنفرانس‌های بین‌المللی (پیم، ۲۰۱۴). بنابراین افراد غیرانگلیسی زبان برای بهبود مهارت‌های زبان انگلیسی خود تحت فشار دائمی هستند، که می‌تواند منبعی از اضطراب و بار عاطفی باشد. وقتی ما فقط به زبان انگلیسی می‌نویسیم، روش خود را برای توصیف روابط بین ایده‌ها محدود می‌کنیم - نوعی از دست دادن که با ایجاد یک "تک فرهنگ معرفت‌شناختی" تشبیه شده است (بنت، ۲۰۱۳). علاوه بر این، محدود کردن بحث‌های علمی جهانی به یک زبان واحد می‌تواند افرادی را که دانش علمی را می‌سازند، به آن دسترسی دارند و به عموم مردم انتقال می‌دهند محدود می‌کند (گاسپاری و دیگران، ۲۰۱۵)، و تک‌زبانگی علمی ممکن است انتشار علم را برای مؤسسات و جوامع غیر انگلیسی زبان کاهش دهد، که می‌تواند دانش جدید را برای افرادی که برای آنها مرتبط‌تر است، مانند کسانی که در نزدیکی سایت‌های مطالعه، رسانه‌های عمومی محلی، و سیاست‌گذاران منطقه‌ای زندگی نمی‌کنند، غیرقابل دسترس کند. در حالی که قطع ارتباط بین علم و جامعه برای هر رشته علمی ناخوشایند است، هزینه آن به ویژه برای علوم کاربردی و رشته‌های مربوط به بحران مانند علوم آب و هوا، اپیدمیولوژی و حفاظت بالا است. در این رشته‌ها انتشار سریع نتایج جدید باعث ایجاد تفاوت اساسی در تصمیمات فوری می‌شود که باید علیرغم شواهد ناقص اتخاذ شوند.

وجود یک زبان جهانی علم واحد ممکن است در حال حاضر برای به اشتراک گذاشتن دانش جدید به طور گسترده و عملی مفید باشد. با این حال، کسانی که هزینه‌های یک زبان واحد را متحمل می‌شوند، با موانع بیشتری نیز مواجه می‌شوند، برای مثال موانعی که با استعمار مرتبط هستند، زیرا زبانی که یک فرد به آن صحبت می‌کند با تاریخ کشور و فرهنگ آن‌ها گره خورده است. بنابراین، حفظ یک زبان واحد در علم بدون ارائه حمایت کافی از افرادی که به آن زبان صحبت نمی‌کنند، به عدم تعادل تاریخی ادامه خواهد داد. تلاش برای ایجاد یک زبان متمرکزتر در دسترس (مانند اسپرانتو) مورد توجه قرار نگرفته است (دایزور و دیگران، ۲۰۱۵). در حالی که انگلیسی ممکن است برخی از مزایای زبانی (مانند دستور زبان نسبتاً ساده و بدون جنسیت) را ارائه دهد، این تنها زبانی نیست که این ویژگی‌ها و ویژگی‌های آن را دارد. تسلط را می‌توان به عوامل تاریخی ذکر شده در بالا و ساختار دستوری نسبتاً ساده آن نسبت داد. بنابراین، بهترین شکل این است که علم از ادغام چندین زبان سود ببرد. علم چندزبانه همچنین با ایجاد سیستم‌های پشتیبانی که می‌تواند انتقال‌های بالقوه آینده را

تسهیل کند، به نفع جامعه خواهد بود، زیرا اگرچه ممکن است بعید به نظر برسد، اما تاریخ نشان داده است که زبان‌های غالب احتمالاً در طول زمان تغییر خواهند کرد. در حال حاضر، پشتیبانی ساختاری کمی برای کمک به غیرانگلیسی‌زبانان برای انتشار و پیشرفت حرفه‌ای به زبان انگلیسی وجود دارد. آمانو و همکاران (۲۰۲۱) برخی نکات عملی را برای غلبه بر موانع زبانی، مانند ترویج فعالیت‌های چندزبانه و همدلی با کسانی که با موانع زبانی مواجه هستند را مطرح کردند.

ترجمه‌ها قبلاً توسط مترجمان انسانی به صورت دستی انجام می‌شد، اما با توجه به محدودیت‌های ذاتی (تخصص موضوعی و هزینه‌های زیاد)، تمرکز از آن زمان به سمت توسعه ترجمه ماشینی معطوف شد. ابزارهای ترجمه ماشینی می‌توانند به دانشمندان کمک کنند تا گام‌های ملموسی در جهت انتشار به چندین زبان از جمله به زبان انگلیسی بردارند. ترجمه ماشینی یا ترجمه خودکار از طریق یک برنامه نرم‌افزاری بدون کمک انسانی انجام می‌شود. ترجمه ماشینی به عنوان یک زیرشاخه از زبان‌شناسی محاسباتی و مهندسی زبان تکامل یافته است. ترجمه ماشینی ایده‌ها و تکنیک‌های خود را از زبان‌شناسی، علوم کامپیوتر، هوش مصنوعی، تئوری ترجمه و آمار می‌گیرد. اولین رویکردهای ترجمه ماشینی از قواعد زبانی برنامه‌ریزی شده و دیکشنری‌های بسیار بزرگ استفاده می‌کردند، اما موفقیت محدودی داشتند زیرا زبان پر از ابهام است و رایانه‌ها به نوع دانش دنیای واقعی و تعاملات اجتماعی که مردم برای تفسیر زبان استفاده می‌کنند، دسترسی نداشتند (راه، ۲۰۲۰). سه رویکرد اصلی برای ترجمه ماشینی وجود دارد: رویکرد مبتنی بر قانون - در اینجا نرم‌افزار مجموعه‌ای از قوانین دستور زبان را برای تکمیل خودکار فرآیند ترجمه به کار می‌گیرد. این قوانین توسط متخصصان زبان انسانی و برنامه‌نویسانی که می‌دانند چگونه قوانین دو زبان را ترسیم کنند، ایجاد شده است. علاوه بر این، این سیستم همچنین به فرهنگ لغت دوزبانه ساخته شده برای این منظور متکی است. در ترجمه ماشینی آماری سیستم از الگوریتم‌های پیش‌بینی استفاده می‌کند تا به رایانه آموزش دهد که چگونه یک متن را از یک زبان به زبان دیگر ترجمه کند. ساخت الگوریتم‌ها (مدل‌های ترجمه آماری) یک فرآیند سریع است، اما پارامترهای ایجاد یک مدل مناسب از تجزیه و تحلیل مجموعه‌های متنی دو زبانه، مجموعه‌ای از نمونه‌های ترجمه خوب ناشی می‌شوند. سپس از مدل آماری برای ترجمه خودکار متن داده شده برای محتمل‌ترین خروجی استفاده می‌شود. مدل‌های ترجمه ماشینی آماری را می‌توان به زیرگروه‌هایی مانند مدل‌های ترجمه مبتنی بر کلمه، مبتنی بر عبارت و مبتنی بر سلسله مراتب سازمان‌دهی کرد. ترجمه ماشینی ترکیبی، ویژگی‌های دو رویکرد فوق را ترکیب می‌کند. ترجمه ماشینی عصبی (NMT) به عنوان جدیدترین روش تکامل

یافته است که بر اساس شبکه‌های عصبی موجود در مغز انسان مدل سازی شده است. اطلاعات از طریق «لایه‌های الگوریتم‌های» متفاوتی می‌گذرد که متن ورودی را برای ترجمه خروجی پردازش می‌کند. نرم‌افزار NMT همچنین از اشتباهات خود درس می‌گیرد، سبک‌های نوشتاری مورد استفاده را انتخاب می‌کند و از شبکه عصبی تکراری (RNN) برای به حداکثر رساندن کیفیت ترجمه استفاده می‌کند. در حالی که هنوز برای ارزیابی تأثیر بالقوه این رویکرد زود است، در حال حاضر بسیار واضح است که این رویکرد خروجی را بهبود می‌بخشد که برای برآورده کردن انتظارات کاربر نهایی کافی به نظر می‌رسد. به دنبال معرفی اینترنت و روند فزاینده تولید متون به شکل دیجیتال، محققان ترجمه ماشینی از رویکردهای زبانی فاصله گرفتند و به سمت ترجمه ماشینی مبتنی بر داده‌ها رفتند که از نقاط قوت رایانه‌ها (مانند تطبیق الگو، محاسبات سریع) استفاده کند. در اوایل قرن بیست و یکم، سیستم‌های ترجمه ماشینی آماری، از جمله ابزارهای آنلاین رایگان اولیه، مانند ترجمه ماشینی گوگل ظاهر شدند. در ترجمه ماشینی آماری، توسعه دهندگان کامپیوتر را با مقادیر زیادی از متون ترجمه شده قبلی تغذیه کردند و سیستم از این مثال‌ها برای محاسبه احتمال ترجمه یک عبارت معین به روشی خاص در متن آینده استفاده کرد (وی، ۲۰۲۰). ابزارهای ترجمه ماشینی آماری خروجی با کیفیت بهتری نسبت به رویکردهای زبانی تولید کردند، اما هنوز جای قابل توجهی برای بهبود وجود داشت. یکی دیگر از رویکردهای مبتنی بر داده، معروف به ترجمه ماشین عصبی، در اواخر سال ۲۰۱۶ ظاهر شد و جهش دیگری را از نظر کیفیت ترجمه ارائه کرد. امروزه، اکثر ابزارهای ترجمه ماشینی از شبکه‌های عصبی مصنوعی در ترکیب با تکنیک‌های مبتنی بر هوش مصنوعی مانند یادگیری ماشین استفاده می‌کنند (فورکادا، ۲۰۱۷). این تکنیک‌ها از توسعه‌دهندگان می‌خواهد که نمونه‌های آموزشی بسیاری از متون منبع اصلی و ترجمه‌های آن‌ها را برای سیستم ترجمه ماشینی ارائه دهند تا سیستم یاد بگیرد. بنابراین، ابزارهای ترجمه با تعداد بیشتری از این مثال‌ها، راحت‌تر به زبان‌ها یا زبان‌های پرکاربرد تنظیم می‌شوند. اگرچه سیستم‌های ترجمه ماشینی عصبی کامل نیستند، اما نسبت به سیستم‌های ترجمه ماشینی قدیمی‌تر که بر رویکردهای زبان‌شناختی یا آماری متکی بودند، نقطه شروع مناسب‌تری ارائه می‌دهند. نتایج سیستم‌های ترجمه ماشینی عصبی را می‌توان برای کسب دانش پایه یا به‌عنوان پیش‌نویس اولیه استفاده کرد که می‌توان آن را بهبود بخشید (به عنوان مثال، برای نوشتن آکادمیک (گولت، ۲۰۲۰). تعداد فزاینده‌ای از مردم از ابزارهای ترجمه ماشینی عصبی به دلیل سهولت استفاده و در دسترس بودن آنلاین رایگان استفاده می‌کنند. مثل دیپل و ترجمه ماشینی گوگل (باوکر، ۲۰۲۱). با این حال، استفاده

از ابزارهای ترجمه ماشینی هنوز نیاز به قضاوت دارد، به همین دلیل است که نیاز به سواد ترجمه ماشینی وجود دارد. فناوری‌های یادگیری ماشینی به کمیت و کیفیت داده‌های آموزشی خود بسیار حساس هستند. برای اینکه سیستم‌های ترجمه ماشینی خوب کار کنند، نه تنها به مقادیر زیادی از متون ترجمه شده قبلی و متن‌های اصلی مربوط به آنها، بلکه به متون با کیفیت خوب که مرتبط با موضوع قانونی هستند نیز نیاز دارند.

به عنوان مثال، زبان مورد استفاده در زمینه‌های تخصصی شامل بسیاری از اصطلاحات و ساختارهای فنی است که بخشی از زبان روزمره نیستند. بنابراین، برای یک سیستم ترجمه ماشینی برای ترجمه دقیق متون در زمینه زیست‌شناسی، باید میلیون‌ها نمونه از متون ترجمه شده قبلی به طور خاص از این حوزه ارائه شود. علاوه بر این، این مثال‌ها باید تمام ترکیب‌های زبانی مورد نظر را پوشش دهند (به عنوان مثال، انگلیسی و فرانسوی، چینی و هندی، انگلیسی و چینی). در برخی موارد، زمانی که برای یک جفت زبان خاص متن‌های ترجمه شده نسبتاً کمی در دسترس است، کمبود داده‌های آموزشی را می‌توان با استفاده از یک زبان رایج به عنوان زبان محوری (مانند ترجمه از اسپانیایی به چینی با استفاده از انگلیسی به عنوان یک واسطه) برطرف کرد. به طور مشابه برای ارتباطات گفتاری، مسایل مورد توجه عمومی بایستی سرعت و استفاده از بسترهای ارتباطی آنلاین را افزایش داد که زیرنویس‌های بسته را به چندین زبان ارائه می‌کنند.

گام‌های روشنی وجود دارد که دانشمندان و توسعه دهندگان ابزار ترجمه ماشینی می‌توانند برای بهبود اجرای فناوری‌ها در ترجمه علمی بردارند. تلاش هماهنگ برای ارائه ترجمه‌های با دسترسی آزاد، تأیید شده توسط انسان و با کیفیت بالا از چکیده‌ها در مجلات علمی، به طور قابل توجهی به تولید داده‌های لازم برای آموزش سیستم‌های ترجمه ماشینی کمک می‌کند. در حال حاضر، ابزارهای ترجمه آنلاین رایگان عمدتاً بر روی داده‌های زبان عمومی به جای اصطلاحات علمی یا زبان تخصصی آموزش می‌بینند. محققان و توسعه دهندگان ابزار می‌توانند در ابزارهای دسترسی باز که سیستم‌های ترجمه ماشینی را برای زمینه‌های تخصصی تحقیق آموزش می‌دهند، همکاری کنند. همزمان به‌طور کلی، می‌توانیم دانشمندان را تشویق کنیم که واژه‌نامه‌های چندزبانه اصطلاحات تخصصی را توسعه دهند یا در آن مشارکت کنند، تا حدی برای کمک به همگام شدن با نسل دائمی اصطلاحات علمی جدید به عنوان مثال، ویکی‌پدیا یک پلتفرم با دسترسی آزاد عالی برای یافتن ترجمه‌های چندزبانه موضوعات فنی و علمی است. با این حال، در حال حاضر توسط چندین رشته علمی کمتر از آن استفاده می‌شود، و چندین زبان با تعداد زیادی گویشور (مانند هندی و ترکی) کمتر در آن مشارکت دارند. از جمله

حوزه‌های پژوهشی مهم ترجمه ماشینی می‌توان به بررسی کارآمدی این نظام‌ها در اجرای دقیق ترجمه متون اشاره نمود. بدیهی است که علاوه بر سرعت، کیفیت ترجمه ماشینی از نکات حائز اهمیت بشمار می‌رود. این افزایش کیفیت در طول زمان به مدد ارزیابی نظام‌های موجود و مشخص شدن نقاط ضعف آنها حاصل شده است. پژوهش در این حوزه با شتاب و وسعت مثال زدنی به پیش می‌رود و جنبه‌های اقتصادی ترجمه ماشینی و نیاز به معیارهای جدید بررسی کارآمدی از حوزه‌های پژوهشی دیگری است که به موارد پیشین اضافه شده است. در نهایت اینکه بازار ترجمه ماشینی به شکلی سریع در حال پیشرفت است و شرکت‌های چند ملیتی در حال برداشتن گام‌هایی برای بومی سازی محتوا به زبان‌های بیشتری هستند. تقاضا برای محتوای مبتنی بر مکان نیز در بخش‌های مختلف صنعتی مانند تجارت الکترونیک، الکترونیک، مسافرت، بازاریابی و توریسم در حال افزایش است. این که فناوری به ارائه اطلاعات به کاربران نهایی به زبان مادری آنها کمک می‌کند، یک واقعیت پذیرفته شده است. غایت این فناوری در واقع نیاز مطلق هر سازمان برای کسب شهرت در تجارت و دستیابی به بازار جهانی ترجمه ماشینی است.

## ۲-۲-۲ علم سنجی و تحلیل هم واژه

علم سنجی شاخه‌ای نسبتاً جدید از علم است که به دنبال سنجش و تجزیه و تحلیل برون‌دادهای محصولات علمی به منظور ساده سازی استراتژی‌های برنامه ریزی سازمان‌ها، ارائه مدیریت مؤثر منابع مالی و انسانی، افزایش کارایی مؤسسات علمی و هدایت آنها در مسیر صحیح است. قدمت این رشته علمی به دهه ۱۹۷۰ برمی‌گردد و به طور گسترده مورد استفاده قرار گرفته است. تا به امروز، روش‌های علم سنجی برای مسائل مختلف تحقیقاتی به کار گرفته شده است. روش‌های علم سنجی می‌توانند برای آشکارسازی چالش‌های مهم پژوهشی آینده به کار گرفته شوند. علم سنجی که گاه کتاب سنجی هم نامیده می‌شود، به طور گسترده در زمینه تجزیه و تحلیل آماری و کمی انتشارات دانشگاهی برای به تصویر کشیدن نقاط داغ، گرایش‌ها، و مشارکت محققان و مجلات و همچنین کشورها به کار گرفته شده است، که به نوبه خود می‌تواند به دانشمندان کمک کند تا به روز باشند.

تحلیل هم واژه به عنوان یکی از روش‌های تحلیلی علم سنجی و نگاشت علم، نقش مهمی را در ایجاد

رهنمودها و ارزیابی روندهای پژوهشی ایفا می‌کند. این میشل کالن بود که در سال ۱۹۸۳ یکی از مهم‌ترین روش‌های کتاب‌سنجی را اختراع کرد یعنی تحلیل هم‌واژه، این تحلیل می‌تواند نقطه داغ یک زمینه معین را با استفاده از هم‌رخدادی واژه‌ها یا عبارات نشان دهد. اصل اساسی این تکنیک بر اساس فراوانی همزمان کلیدواژه‌های موضوعی و عبارات در یک مقاله یا پیکره متن است. این بسامد، سپس می‌تواند برای انجام تحلیل و خوشه‌بندی سلسله‌مراتبی و سپس تجزیه و تحلیل ارتباط این کلمات یا عبارات پیاده‌سازی شود. از آن زمان، محققان دانشگاهی از تحلیل هم‌کلمه برای ترسیم ساختار کتاب‌سنجی حوزه‌های مختلف از جمله خلاقیت استفاده کرده‌اند (ژانگ، و همکاران، ۲۰۱۵). تحلیل هم‌واژه روشی موثر برای تحلیل محتوا و متن کاوی نیز در نظر گرفته شده است (چاتر، ۲۰۱۵). یکی از مزایای کلیدی آن این بود که ساختار مفهومی یک رشته را بدون نیاز به مراجعه به متن کامل آشکار کرد (رومو-فرناندز و همکاران، ۲۰۱۳). این باور وجود دارد که کلمات کلیدی به دقت توسط نویسندگان انتخاب شده و محتوای سند را به طور دقیق نشان می‌دهد (فنگ و همکاران، ۲۰۱۷). تحلیل هم‌واژه را می‌توان برای تعیین کمیت پیوندهای بین موضوعات تحقیق در یک رشته علمی استفاده کرد. در سال‌های اخیر، این تکنیک در زمینه‌های موضوعی مختلف به کار گرفته شده است و توجه بسیاری از محققان را به خود جلب کرده است. فرض اصلی این تکنیک این است که کلمات کلیدی یک مقاله توصیف قابل قبولی از محتوای آن ارائه می‌دهند. تحلیل هم‌واژه رابطه بین ایده‌ها و مفاهیم را تعیین می‌کند. ظهور دو کلمه در یک متن علمی نشان دهنده رابطه بین موضوعاتی است که این دو کلمه به آنها اشاره می‌کنند. این روش در مطالعات علم‌سنجی زمینه‌های مختلف و برای بسیاری از موضوعات مورد بهره‌برداری قرار گرفته است. در سال‌های اخیر، روش‌ها و رویکردهای جدیدی توسط محققان برای تحلیل هم‌واژه پیشنهاد شده است. ژانگ و همکاران (۲۰۱۷) از تحلیل هم‌واژه در کنار تحلیل مقیاس‌بندی چندبعدی و تحلیل شبکه‌های اجتماعی برای تعیین موضوعات و فرصت‌های تحقیقاتی مهم و به تصویر کشیدن روند تکاملی حوزه خلاقیت استفاده کرد. وانگ و همکاران به جای استفاده از تکنیک مرسوم تکرار کلمات، روش جدیدی را توسعه داد. این روش شبکه‌های همزمان پویا با تمرکز بر پیوند بین کلمات کلیدی بود. آنها نرم‌افزاری به نام NEViewer برای ادغام تکنیک‌ها، الگوریتم‌ها و اندازه‌گیری‌ها در

روش‌های علم سنجی مختلف از جمله شبکه‌های هم‌رخدادی، هم‌نویسندگی و هم‌استناد ایجاد کردند. آنها از نمودارهای آبرفتی و رنگ آمیزی استفاده کردند. سیترز و همکاران روشی را برای شناسایی خوشه‌های موضوعی پیشنهاد کردند. این روش شامل دو مرحله بود. مرحله اول به شناسایی مقاله اختصاص داشت، در حالی که واژه پردازی در مرحله دوم انجام شد که بر اساس آن روند شناسایی موضوعات مهم می‌تواند ارائه شود. ژانگ و همکاران (۲۰۱۵) تئوری‌های شبکه‌های پیچیده را برای تحقق بخشیدن به یک شبکه همزمان پویا در زمینه مطالعاتی خود، یعنی پژوهش شهرنشینی چین به کار گرفتند. آنها همبستگی درجه، بردار ویژه و مرکزیت بین شبکه‌ها را در دوره‌های مختلف بررسی کردند. در یافته‌های این پژوهش مشخص شد که درجه بینابینی و درجه مرکزیت همبستگی مثبت دارند. فنگ و همکاران (۲۰۱۷) تحلیل هم‌واژه را با فاصله معنایی ترکیب کردند تا دقت نتایج تجزیه و تحلیل را بهبود بخشند زیرا معتقد بودند این دو مفهوم دارای یک ارتباط معنایی خاص هستند. ژو و همکاران (۲۰۱۲) روش اندازه‌گیری معنایی را در تحلیل هم‌واژه ارائه کردند تا به ارتباط هم‌رخدادی واژه‌ها با کیفیت‌های مختلف و عدم وجود معنانشناسی در محاسبات همبستگی پرداخته شود. آنها وزنی را برای هر ویژگی به دست آمده در نظر گرفتند، سپس کلیدواژه‌های اصلی را استخراج کردند. در نهایت، پس از تشکیل خوشه‌ها و تبدیل به بردارهای توزیع ارزش، همبستگی معنایی بین کلمات کلیدی محاسبه شد. کاتسورا و اونو تکنیکی را بر اساس شبکه هم‌واژه پویا پیشنهاد کردند و شبکه‌های هم‌واژه پویا پراکنده را به دست آوردند. آنها هر شبکه را با موضوعات پراکنده در یک دوره مربوطه دیداری سازی کردند.

### ۳-۲-۲ سنجش و ارزیابی و ترجمه ماشینی

پژوهش در زمینه ترجمه ماشینی به شیوه‌های مختلف انجام می‌شود. بررسی کارآمدی سیستم در آرایه ترجمه‌های دقیق و مناسب یکی از این شیوه‌ها است. دیگر روش ارزیابی کیفیت ترجمه ماشینی استفاده از نظرات کاربران است. استفاده از روش علم سنجی در ارزیابی انتشارات حوزه ترجمه ماشینی و حوزه‌های مرتبط دیگر روش سنجش است که می‌توان بکار برد. پژوهش در این زمینه بسیار نادر است. در ادامه به یکی از این موارد اشاره می‌شود. ووس و ژایو (۲۰۰۵) در

پژوهشی روند تحقیقات ترجمه ماشینی توسط ۶۸۳ سازمان در ۹۳ کشور را بررسی کردند. توزیع تحقیقات در بین سازمان‌های مشارکت کننده به شدت نامتجانس مشاهده شد. اکثر سازمان‌های شرکت کننده (۵۳۹) هر کدام ۱-۱۰ مقاله، و ۱۰۷ سازمان دیگر هر کدام ۱۱-۳۰ مقاله، ۲۰ سازمان هر کدام ۳۱-۵۰ مقاله، ۱۳ سازمان هر کدام ۵۱-۱۰۰ مقاله و ۴ سازمان هر کدام ۱۰۱-۱۵۴ مقاله منتشر کرده بودند. ۵۰ سازمان مولد برتر در تحقیقات ترجمه ماشینی ۴۷,۳۸ درصد از سهم انتشار جهانی و ۶۱,۳۹ درصد از سهم استناد جهانی در طول سال‌های ۲۰۱۶-۲۰۰۷ سهم داشتند. هفده مورد از این سازمان‌ها خروجی انتشارات خود را بالاتر از میانگین گروه ۴۹,۱ ثبت کردند: دانشگاه شهر دوبلین، ایرلند (۱۵۳ مقاله)، دانشگاه کارنگی ملون، ایالات متحده (۱۱۹ مقاله)، موسسه ملی فناوری اطلاعات و ارتباطات ژاپن، (۱۱۵ مقاله)، موسسه فناوری هاربین، چین (۹۹ مقاله) منتشر کرده بودند. پژوهش حاضر به دلیل کمبود پیشینه در این حوزه اهمیت زیادی دارد.

### ۳,۲- مرور پیشینه‌های بین‌المللی

پس از آنکه اسمال (۱۹۷۳) برای نخستین بار از نگاشت هم‌استنادی برای مطالعه و تحلیل تکامل حوزه‌های علمی استفاده کرد، توجه جامعه پژوهشگران به توانایی‌های این ابزار پژوهشی معطوف گردید. پژوهشگران فرانسوی نظیر کالون و دیگران (۱۹۸۳) با ارائه رویکرد دیگری در تحلیل‌های کمی علم‌سنجی، از واژه‌ها به جای استناد به عنوان واحدهای تحلیل استفاده کرده و نگاشت هم‌واژه را مطرح کردند.

به عقیده آنها، واژه‌ها اجزای اصلی تشکیل دهنده مبنای موضوعی مقالات هستند. با اندازه‌گیری شدت ارتباط بین آنها که بر اساس محاسبه هم‌رخدادی صورت می‌گیرد، می‌توان به تحلیل گرایش‌های پژوهشی و موضوعی در حوزه‌های مختلف دست یافت.



این رویکرد سپس در کارهای کالون و دیگران (۱۹۸۶) و لیدزدورف (۱۹۸۹) به بلوغ نسبی رسید و به عنوان جایگزینی برای نگاشت هم استنادی مطرح شد. از آن زمان محققان بسیاری از این فنون برای تحلیل و نگاشت کتابسنجی حوزه‌های مختلف استفاده کرده اند. از جمله آنها می‌توان به دینگ و دیگران (۲۰۰۱) در تحلیل و نگاشت حوزه بازیابی اطلاعات اشاره کرد. وی ساختار علمی حوزه بازیابی اطلاعات در طی سال‌های ۱۹۸۷-۱۹۹۷ را با تحلیل هم واژه بررسی نمود تا الگوها و گرایش‌ها در این حوزه با محاسبه شدت ارتباط بین اصطلاح‌هایی که بازنمون انتشارات مرتبط هستند مشخص شود. وی از داده‌های تامسون رویترز استفاده نمود. کلید واژه‌ها از عناوین و چکیده استخراج شدند. آنگاه با استفاده از ابزار کنترل واژگان استاندارد سازی واژگان انجام شد. یافته‌های این پژوهش، پویایی این رشته در سالیان مورد بررسی و تغییرات جدید در حوزه‌های اصلی پژوهشی آن را نشان داد.

بورنر (۲۰۰۵) نیز با بررسی مجموعه مقالات کنفرانس **PNAS** تحلیل هم واژه و واژه‌های پر انتشار را در خصوص ده درصد از پراستنادترین مقالات مجموعه مذکور انجام داد. در پژوهش حاضر، برای مطالعه واژه‌های پراستناد که در یک دوره زمانی ناگهان به شدت از سوی جامعه علمی استفاده شده اند از الگوریتم کلینبرگ که در پژوهش بورنر هم مورد استفاده قرار گرفته بود، استفاده می‌شود.

در سطح بین‌المللی نیز چنین پژوهش‌هایی کمتر انجام شده است. تنها موردی که یافت شد اثر ووب و ژائو ۲۰۰۵ است که به بررسی علم سنجی حوزه ترجمه ماشینی پرداخت. یافته‌های آنها نشان داد که ترجمه ماشینی به شاخص‌هایی علاوه بر کارآمدی احتیاج دارد. به همین دلیل اولین گام‌ها را در تحلیل علم سنجی ترجمه ماشینی برداشتند. تحلیل آنها منحصر به استفاده از مجموعه مقالات یک کنفرانس آمریکایی در حوزه زبانشناسی و بررسی آن با شاخص‌های علم سنجی بود. همانگونه که مشخص است تحلیل آنها به دلیل محدودیت پیکره مورد بررسی قابل اعتماد و تعمیم نمی‌باشد.

پیشینه‌های مربوط به مطالعات کتاب سنجی در مورد تحقیق ترجمه ماشینی و حوزه‌های مرتبط در این زمینه به چند انتشار محدود است. ووس و ژو (۲۰۱۶) تحقیقات ترجمه ماشینی را با استفاده از شاخص‌های انتشارات تحلیل کردند. یافته‌های پژوهش آنها نشان داد که در انجام ترجمه ماشینی می‌توان از رویکردهایی چون پیکره موازی، واژه نامه‌های الکترونیکی و فنون زبانشناسی رایانه ای بهره گرفت اما کارآمدترین روش همان استفاده از پیکره‌های موازی است. دونگ و چن (۲۰۱۶) با استفاده از پایگاه داده **Web of Science**، نشریات را طی سال‌های ۲۰۰۰ تا ۲۰۱۵ تجزیه و تحلیل

کردند. نویسندگان وضعیت فعلی تحقیق در زمینه ترجمه ماشینی در این زمینه را از نظر گرایش‌های انتشار، توزیع‌های جغرافیایی، ادبیات اصلی و حوزه‌های تحقیقاتی متمایز ترجمه ماشینی گزارش کردند: مطالعات ترجمه نظری، آموزش ترجمه و تفسیر و مطالعات ترجمه توصیفی مهم‌ترین گرایش‌ها مطرح شده بر اساس یافته‌های آنها بود. در پژوهشی دیگر، گیل (۲۰۱۴) تحقیقات کتاب‌سنجی در مطالعات ترجمه در چین و پیشرفت‌های اخیر آن را دنبال کرد. یافته‌های وی نشان داد یکی از مهم‌ترین چالش‌های ترجمه ماشینی به زبان چینی، شیوه نوشتار آنها است که ناشی از الفبای چینی است که بصورت به هم چسبیده و فاقد فاصله است، وی راه حلی برای این موضوع ارائه کرد. زانتین و همکاران (۲۰۰۵) چگونگی تعریف زیرشاخه‌های مطالعات ترجمه و نحوه علایق و کانون‌های پژوهشی آن را مطالعه و دریافتند این موارد به تدریج در طول زمان تغییر کرده‌اند.

ترجمه ماشینی استفاده از سیستم‌های رایانه‌ای برای ترجمه متن است. دو زبان منبع و هدف در این فرایند دخیل می‌باشد. هدف این فرایند آن است که مشارکت انسانی در آن به حداقل برسد. رویکردهای مختلفی در این حوزه وجود دارد که جدیدترین و موفق‌ترین آنها شیوه ترجمه ماشینی آماری است. در این رویکرد برای ترجمه از پیکره‌های متنی موازی استفاده می‌شود. این شیوه، برخلاف دیگر رویکردها چندان وابستگی به دانش انسانی ندارد.

امروزه در بسیاری از رشته‌های علمی از تحقیقات علم سنجی برای شناسایی موثرترین مولفین، مقالات و نشریات یک حوزه استفاده شده است. در رشته زبان‌شناسی رایانه‌ای و حوزه ترجمه ماشینی اما چنین پژوهش‌هایی کمتر انجام شده است. پژوهشی با عنوان از زبان‌شناسی رایانه‌ای تا تاریخ نگاری الگوریتمیک (لیانگ و ژو، ۲۰۱۶) یکی از این معدود کارها است. در این پژوهش برخی مشکلات ترجمه ماشینی ذکر شده است. ضمناً بر اهمیت نگاشت معنایی واژه‌ها در پژوهش‌های علم سنجی تأکید ویژه شده است.

در پژوهشی دیگر ووس و ژو (۲۰۰۵) به بررسی علم سنجی حوزه ترجمه ماشینی پرداختند. روش تحلیل کمی و تحلیل استنادی مقالات این حوزه نشان داد که ترجمه ماشینی به شاخص‌هایی علاوه بر کارآمدی احتیاج دارد. این پژوهش از جمله اولین گام‌ها در تحلیل علم سنجی ترجمه ماشینی است. در آن تحلیل مجموعه مقالات یک کنفرانس آمریکایی در حوزه زبان‌شناسی با شاخص‌های علم سنجی تحلیل شد. همانگونه که مشخص است این تحلیل به دلیل محدودیت پیکره مورد بررسی قابل اعتماد و تعمیم نمی‌باشد.

## ۴,۲- مرور پیشینه‌های پژوهش‌های ملی

با بررسی پژوهش‌های داخل کشور، پژوهشی که در آن از روش علم سنجی برای بررسی تحقیقات ترجمه ماشینی استفاده شده باشد مشاهده نگردید. اصولاً استفاده از تحقیقات علم سنجی در حوزه زبانشناسی رایانه‌ای و پردازش زبان طبیعی و زیر شاخه‌های آن چندان در ایران انجام نشده است. از موارد محدود در این زمینه می‌توان به پژوهش‌های زیر اشاره کرد:

فلاحتی، احمدی نسب و خانی (۱۳۹۷) به بررسی روند مطالعات ترجمه در ایران پرداختند. آنها با بررسی نشریات رتبه دار وزارت علوم تلاش نمودند تا حوزه‌های پژوهشی برتر و کمبودهای تحقیقاتی در این زمینه را مشخص سازند. در نهایت موضوعاتی چون اصول و روش ترجمه، آموزش ترجمه و ترجمه متون مذهبی از جمله مهم‌ترین حوزه‌های پژوهشی در این مجموعه شناخته شد. تفاوت پژوهش حاضر با این پژوهش در گستره بین‌المللی آن و استفاده از روش تحلیل علم سنجی و دیداری سازی حوزه‌های دانش است.

قلی فامیان (۱۳۹۳) در پژوهشی دیگر به بررسی مجموعه مقالات دو دهه اخیر کنفرانس‌های زبانشناسی با روش علم سنجی پرداخت. وی با بررسی مقالات هفت سمینار زبان‌شناسی داخلی با معیارها و شاخص‌های علم سنجی، نتیجه گرفت که در طول زمان چه حوزه‌هایی بیشتر مورد استقبال قرار گرفته، موثرترین مقالات و مولفین و دانشگاه‌ها کدام بوده و افزایش سهم زنان مولف در تاریخ چقدر افزایش داشته است.

## ۵,۲- استنتاج از پیشینه پژوهش

بررسی پیشینه‌های موجود نشان داد که علی‌رغم انجام تحقیقاتی اندک در زمینه علم سنجی ترجمه ماشینی، اما این حوزه دارای پتانسیل بالایی برای انجام تحقیقات جدید است. تحقیقاتی نظیر تحلیل‌های هم‌استنادی برای شناخت جنبه‌های تحقیق و همکاری بین مولفان و کشورها در زمینه مورد بحث، اما تحقیق حاضر تنها تحقیقی است که با روش همواژه انجام شده و از این لحاظ دارای نوآوری در این حوزه در زبان فارسی است.

## فصل سوم

### روش شناسی پژوهش



### ۳-۱. مقدمه

در این فصل روش پژوهش شرح داده می شود. ابتدا نوع پژوهش، آنگاه جامعه پژوهش، روش جمع آوری داده ها و در نهایت شیوه تحلیل آنها شرح داده می شود.

### ۳-۲. روش پژوهش

این پژوهش کاربردی است که در آن از روش علم سنجی استفاده می شود. علم سنجی گونه ای پژوهش کمی است که با استفاده از روش های آماری به بررسی برون داده های علمی و سنجش الگوی تکامل رشته های مختلف می پردازد. روش علم سنجی به دو بخش تحلیل عملکرد و دیداری سازی حوزه های موضوعی تقسیم می گردد (کوبو، ۲۰۱۱). در این پژوهش هر دو بخش مورد استفاده قرار گرفته است. در پاسخ به سوالاتی که عملکرد کمی را مورد بررسی قرار می داد از روش تحلیل عملکرد و در بخشی که به بررسی ساختار حوزه ترجمه ماشینی از طریق نقشه های علمی می پرداخت از روش دیداری سازی حوزه های موضوعی بهره گرفته شد. برای دیداری سازی شبکه پژوهش های ترجمه ماشینی از تحلیل هم واژه استفاده شد. در مورد تحلیل هم واژه در فصل دوم قسمت مبانی نظری به تفصیل صحبت شد.

### ۳-۳. نوع پژوهش

این پژوهش کاربردی بوده و در آن از روش علم سنجی استفاده شده است. در جهت تحلیل موضوعی و بررسی ساختار و تکامل انتشارات حوزه ترجمه ماشینی نیز از روش تحلیل هم واژه که نوعی تحلیل محتوا است بهره گرفته شده است.

### ۳-۴. جامعه پژوهش

جامعه پژوهش را کلیه مقالات منتشره حوزه ترجمه ماشینی در سطح بین المللی در سالهای ۲۰۰۶ تا ۲۰۱۵

تشکیل می دهد. این پیکره با تحلیل کتابسنجی مورد بررسی قرار گرفته و اطلاعات مورد نیاز از آن استخراج شده است. در این پژوهش پیشینه‌های موجود در پایگاه استنادی اسکوپوس که در حوزه تحقیقات ترجمه ماشینی انجام شده است استخراج شد. پایگاه اسکوپوس عظیم ترین پایگاه استنادی جهان است و از بدو پیدایش آن روند استفاده از این پایگاه صعودی و رو به پیشرفت بوده است. در این پژوهش از نمونه گیری استفاده نشد تا نمای کلی پژوهش در حوزه ترجمه ماشینی مشخص شود.

### ۵-۳ روش گردآوری داده ها

در مرحله استخراج داده‌ها، با جستجوی عبارت "machine translation" در فیلد عنوان، چکیده و کلیدواژه پایگاه استنادی اسکوپوس داده‌های مربوط استخراج گردید. تعداد کل داده‌های بازیابی شده بالغ بر ۴۴۰۸ رکورد بود که از این میزان ۳۱۷۲ مقاله کنفرانس و ۹۵۹ مقاله نشریه بود که این نشان دهنده ماهیت فنی و تخصصی تحقیقات در این زمینه است. در جهت جامعیت پیکره پژوهش از هر دو نوع مقالات کنفرانس و مقالات نشریه استفاده شد. در مرحله تحلیل عملکرد، موثرترین مولفان و دانشگاه‌هایی که در زمینه تحقیقات حوزه ترجمه ماشینی مطرح هستند شناسایی شدند. سپس واژه‌های کلیدی انتشارات حوزه ترجمه ماشینی از عناوین و چکیده مقالات پیکره پژوهش استخراج گردید. این واژه‌ها برای تحلیل هم واژه مورد استفاده قرار گرفت. در مرحله بعد پیش پردازش کلید واژه‌ها برای پاکسازی داده‌ها انجام شد. این عملیات شامل مراحل مختلفی بود. ابتدا یکدست سازی جمع و مفرد کلیدواژه‌ها انجام شد. در این مرحله واژه‌هایی که شکل مفرد آنها ارجحیت داشت، به شکل مفرد آورده شدند. آنگاه علائم اختصاری، به شکل کامل نوشته شدند البته در برخی موارد، شناسایی فرم کامل به سادگی امکان پذیر نبود و نیاز به جستجو در زمینه تخصصی واژه داشت. واژه‌های عمومی مانند مقاله حذف شد تا در تحلیل به عنوان واژه کلیدی بشمار نیاید. جهت تولید ماتریس هم رخدادی واژه‌ها که نشان دهنده گرایشات موضوعی حوزه ترجمه ماشینی است از نرم افزار ووسویور استفاده شد. این نرم افزار از الگوریتم دیداری سازی



مشابهت‌ها برای تشکیل شبکه هم رخدادی واژه‌ها استفاده می‌کند. با تشکیل شبکه واژه‌ها از روش‌های تحلیل شبکه، نگاره‌های تولید شده تفسیر شد. بر این اساس خوشه‌های موضوعی تشکیل شده و روابط بین آنها به عنوان بازنمون متن کاوی پیکره پژوهش معرفی شد. این خوشه‌ها نشان دهنده گرایشات پژوهشی برتر حوزه ترجمه ماشینی است که در بخش بعد به تفصیل در موردشان صحبت می‌شود.

## فصل چهارم

### تجزیه و تحلیل داده‌ها

## ۴-۱. مقدمه

در این فصل نتایج تجزیه و تحلیل داده ها ارائه خواهد شد. علاوه بر پاسخ سوالات ارائه شده در فصل یک، نقشه های علمی مربوط نیز به نمایش در آمده است.

## ۴-۲ سوال اول پژوهش: مولفین پر تولید حوزه ترجمه ماشینی چه کسانی هستند؟

در جدول ۱ لیست ده نفر از برترین مولفین که در سال های ۲۰۰۵ تا ۲۰۱۶ بیشترین مقالات را در حوزه ترجمه ماشینی منتشر کرده اند مشاهده می شود.

جدول ۱: مولفان پر تولید حوزه ترجمه ماشینی

رتبه	نام مولف	تعداد مقاله	نام دانشگاه یا موسسه پژوهشی	کشور
۱	Hermann Ney	۱۱۹	Rheinisch-Westfälische Technische Hochschule Aachen	آلمان
2	Eiichiro Sumita	95	Japan National Institute of Information and Communications Technology	ژاپن
3	Andy Way	90	Dublin City University	ایرلند
4	Francisco Casacuberta	71	Universitat Politècnica de València	اسپانیا
5	Qun Liu	65	Info Huawei Noah's Ark Lab	هنگ کنگ
6	Philipp Koehn	53	Meta AI	آمریکا
7	Marcello Federico	48	Amazon.com, Inc.,	آمریکا
8	Stephan Vogel	48	Translation and Interpreting Institute, Doha	قطر
9	Tiejun Zhao	48	Info Harbin Institute of Technology	چین

چین	Harbin Institute of Technology	47	Li, Sheng	10
-----	--------------------------------	----	-----------	----

بر اساس جدول فوق پرتالیف ترین مولفان این حوزه به ترتیب نی ۱۱۹، سومیتا ۹۵ و وی ۹۰ مقاله را منتشر کرده اند. سایر محققان برتر به ترتیب مشخص شده اند. باید به خاطر داشت که حوزه ترجمه ماشینی نسبتاً جوان است و انتظار نمی‌رود مولفان پرکار آن همچون سایر رشته‌های پر قدمت باشند. این فهرست می‌تواند به خوانندگان کمک کند تا برجسته‌ترین محققان در این زمینه را شناسایی کنند.

### ۳-۴ سوال دوم پژوهش: دانشگاه‌ها و کشورهای برتر در انتشارات حوزه ترجمه ماشینی در سال های ۲۰۰۵ تا ۲۰۱۶ کدامند؟

یافته های پژوهش در خصوص برترین دانشگاه هایی که در دوره زمانی این پژوهش بیشترین انتشارات مربوط را داشته اند در جدول زیر نشان داده شده است.

جدول ۲: دانشگاه‌های برتر حوزه ترجمه ماشینی

رتبه	نام دانشگاه یا موسسه پژوهشی	کشور	تعداد مقاله
۱	Dublin City University	ایرلند	۱۸۴
۲	Carnegie Mellon University	آمریکا	۱۶۴
۳	Japan National Institute of Information and Communications Technology	ژاپن	۱۵۱
۴	Chinese Academy of Sciences	چین	۱۴۹
۵	Rheinisch-Westfälische Technische Hochschule Aachen	آلمان	۱۳۰
۶	Harbin Institute of Technology	چین	۱۰۶
۷	IBM Thomas J. Watson Research Center	آمریکا	۱۰۴

۱۰۴	اسکاتلند	The University of Edinburgh	۸
۱۰۲	اسپانیا	Universitat Politècnica de València	۹
۹۷	آمریکا	Microsoft Research	۱۰

در جدول ۲ ده دانشگاه برتر نشان داده شد. همانگونه که در این جدول مشخص است دانشگاه شهر دوبلین با ۱۸۴ مقاله، کارنگی ملون با ۱۶۴ مقاله و موسسه اطلاعات ملی ژاپن با ۱۵۱ مقاله پرتولیدترین دانشگاه‌های این حوزه اند. حضور پرتعداد دانشگاه‌های کشورهای صنعتی و پیشرفته در این لیست نشان دهنده ماهیت فنی و نوآورانه این حوزه است.

جدول ۳: کشورهای برتر در انتشارات ترجمه ماشینی

ردیف	کشور	تعداد مقاله
۱	United States	۱۲۰۱
۲	China	۷۷۸
۳	Japan	۵۰۴
۴	Germany	۳۹۰
۵	Spain	۳۷۵
۶	India	۳۲۲
۷	United Kingdom	۳۱۴
۸	France	۲۷۵
۹	Ireland	۲۱۵

۱۶۷	Canada	۱۰
-----	--------	----

یافته های جدول ۳ ده کشوری که بیشترین انتشارات حوزه ترجمه ماشینی را دارند نشان می دهد. در رتبه اول ایالات متحده با ۱۲۰۱ مقاله، رتبه دوم چین با ۷۷۸ مقاله و رتبه سوم ژاپن با ۵۰۴ مقاله قرار دارند. پیشروی چین و ژاپن تا رتبه های دوم و سوم این جدول مبین نفوذ زبان چینی و ژاپنی میان زبان های برتر علم و نیاز به ترجمه آنها به زبان انگلیسی و سایر زبان های پرکاربرد است. همه کشورهای جدول فوق از جمله کشورهای پیشرفته در علوم و فناوری هستند. ماهیت حوزه ترجمه ماشینی نیز به گونه ای است که سرامدی در آن نیازمند بنیان قوی دانش و فناوری است.

#### ۴-۵ سوال سوم پژوهش: توزیع موضوعی انتشارات ترجمه ماشینی چگونه است؟

یافته های پژوهش در خصوص توزیع موضوعی انتشارات ترجمه ماشینی در جدول ۴ مشخص شده است.

جدول ۴: توزیع موضوعی انتشارات ترجمه ماشینی

ردیف	حوزه موضوعی	تعداد مقاله
۱	علوم رایانه ای	۴۲۵۳
۲	هنر و علوم انسانی	۲۲۴۷
۳	علوم اجتماعی	۱۸۳۵
۴	ریاضیات	۸۴۰
۵	مهندسی	۵۸۵
۶	علوم تصمیم گیری	۶۸

۵۷	فیزیک و نجوم	۷
۴۴	حرفه های سلامتی	۸
۴۰	تجارت،مدیریت و حسابداری	۹
۳۴	پزشکی	۱۰

با توجه به یافته های تحلیل موضوعی انتشارات ترجمه ماشینی، حوزه های علوم رایانه ای، علوم انسانی و هنر و علوم اجتماعی بیشترین مشارکت را در تحقیقات ترجمه ماشینی داشته اند. یافته های این سوال قابل پیش بینی است. از آنجا که ترجمه ماشینی به دانش زیاد در حوزه علوم رایانه ای نیاز دارد، بدیهی است که این حوزه در راس سایر موضوعات قرار گیرد. ضمن آنکه چون ترجمه خاستگاهی در علوم انسانی دارد رتبه هنر و علوم انسانی نیز مورد قبول واقع می شود.

**۴-۶ سوال چهارم پژوهش: نقشه های علمی در خصوص ساختار و تکامل رشته ترجمه ماشینی چه اطلاعاتی ارائه می کنند؟**



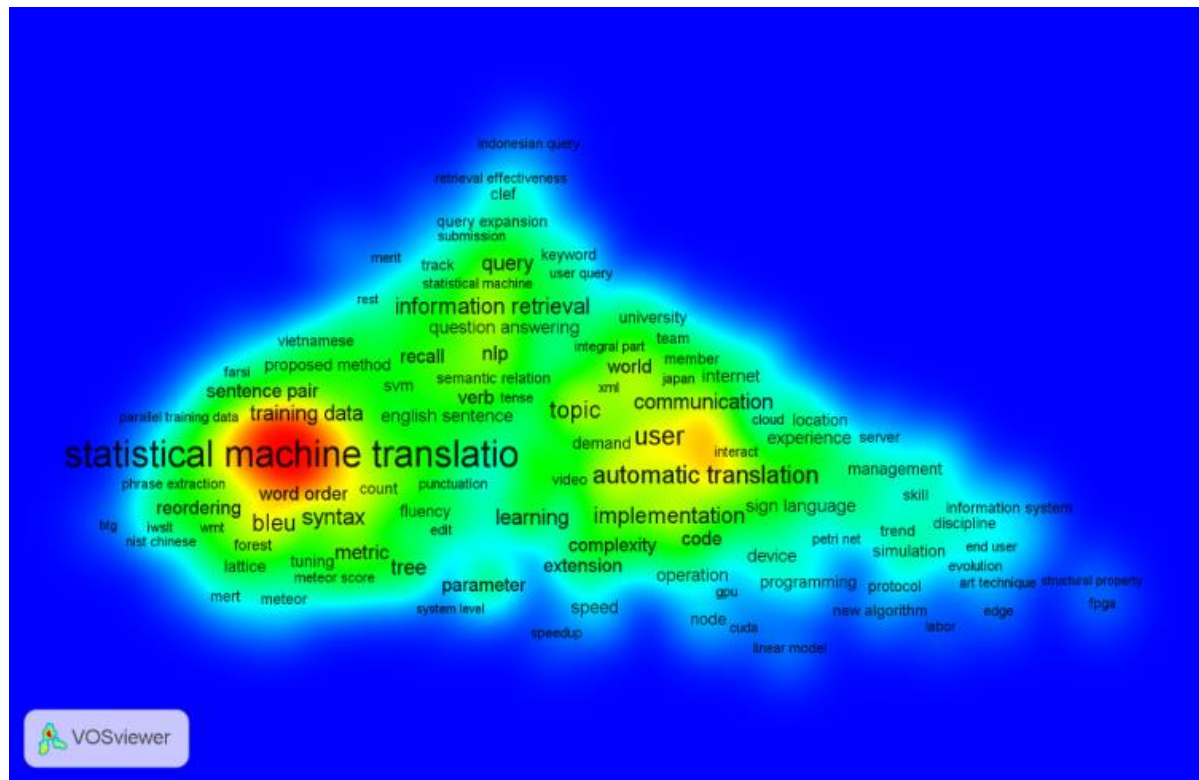


کشیده شده است. کلمات کلیدی که از نظر محتوا مشابه هم هستند در یک خوشه گروه بندی شده اند. اندازه دایره‌ها، نشان دهنده فراوانی کلمه کلیدی و نزدیکی گره‌ها به هم نشان دهنده شدت ارتباط آنها با هم است. همانطور که شکل نشان می‌دهد، همه خوشه‌ها به هم پیوسته بودند و روابط قوی بین آنها وجود داشت. این نشان دهنده وابستگی متقابل بالای حوزه‌های مختلف تحقیقات ترجمه ماشینی است. خوشه موضوعی با نام **ترجمه ماشینی آماری** با بیشترین تناوب و تراکم (سمت چپ شکل) نشان می‌دهد که رویکرد آماری در طراحی ماشین‌های ترجمه در حال حاضر رویکرد اصلی در تحقیقات این حوزه است. اگر چه وجود واژه‌هایی چون نحو و نظم واژگان و درخت واژگان در این خوشه نشان دهنده یکپارچگی این تحقیقات با اصول زبانشناسی رایانشی است. خوشه موضوعی که در وسط و رو به بالای شکل قرار دارد خوشه **بازیابی اطلاعات** است. وجود این خوشه در این شکل بیهوده نیست زیرا بازیابی اطلاعات بویژه در مبحث بازیابی بین زبانی وابستگی مستقیم با ترجمه ماشینی دارد. وجود اصطلاحاتی نظیر پرس و جو، تاپیک و سیستم پرسش و پاسخ موید شکل‌گیری هوشمندانه این خوشه و موفقیت رویکرد استفاده شده در این پژوهش است.

خوشه راست بالای این شکل با نام **ترجمه خودکار** و دارا بودن اصطلاحاتی نظیر کاربر و ارتباطات، نشان دهنده اهمیت ترجمه ماشینی در تعاملات انسانی و ارتباطات بین‌المللی است. در واقع این خوشه وجه علوم انسانی و جامعه‌شناسی را در تحقیقات این حوزه نمایندگی می‌کند.

خوشه پایین راست در این شبکه برخلاف خوشه‌های دیگر از تراکم درونی خاصی برخوردار نیست. بر اساس واژه‌های موضوعی موجود در این خوشه آن را خوشه **طراحی و ارزیابی ترجمه ماشینی** نامگذاری می‌کنیم. بطور کلی شکل حاضر نشان دهنده نحوه تکامل تحقیقاتی رشته بوده و نقشه راه تحقیقات آینده در این حوزه را نیز مشخص می‌کند. نحوه قرار گرفتن این خوشه‌ها که نشان دهنده اشتراک دانش بین آنها از طریق وارد و صادر شدن دانش بین آنهاست ساختار علمی این حوزه را نشان می‌دهد.

در شکل ۲ نمای تراکمی شبکه هم‌واژه ترجمه ماشینی نشان داده شده است.



شکل ۲ نگاره تراکمی حوزه‌های داغ ترجمه ماشینی.

نگاشت کلمات کلیدی از نظر چگالی نیز می‌تواند آموزنده باشد. در نقشه تراکمی که در شکل ۲ دیده می‌شود، هر چه کلمه کلیدی به نواحی قرمز رنگ نزدیک‌تر باشد، رخداد کلمه کلیدی بالاتر است. بنابراین کلمات و عباراتی چون داده‌های آموزشی، نظم‌واژه‌ها، نیاز کاربر، شمارش واژه‌ها و سایر کلماتی که در آن نواحی قرار گرفته‌اند، در سال‌های اخیر بسیار مورد توجه محققان قرار گرفته و عملاً اشباع شده است. در مقابل، کلمات کلیدی مانند پیچیدگی، سنجش، فعل، بازیافت، زبان نمادی و یادگیری توجه نسبتاً کمی را به خود جلب کردند. این یافته می‌تواند محققان را در مورد مؤثرترین حوزه‌هایی که باید روی آنها تمرکز کنند، آگاه کند.

## فصل پنجم

### نتیجه گیری و پیشنهادها



## ۵-۱. مقدمه

در این فصل بحث و نتیجه گیری بر اساس یافته های پژوهش که در فصل چهارم ارائه شد انجام گرفته و سپس پیشنهادهایی برای پژوهش های آتی ارائه می شود.

## ۵-۲. بحث و نتیجه گیری

در مطالعه حاضر انتشارات حوزه ترجمه ماشینی در سالهای ۲۰۰۶ تا ۲۰۱۵ که در پایگاه استنادی اسکوپوس نمایه شده است و بالغ بر ۵۰۰۱ پیشینه است به عنوان پیکره پژوهش مورد بررسی قرار گرفت. از این تعداد ۲۹۸۹ مورد مقاله کنفرانس و ۱۰۸۵ پیشینه از مقالات چاپ شده در نشریات علمی و مابقی از سایر انواع مدارک است. این مطالعه با هدف بررسی ساختار علمی و تکامل حوزه ترجمه ماشینی انجام شده است. در اینجا با مرور یافته های مربوط به سوالات پژوهش به تجزیه و تحلیل آنها پرداخته و گاه با نتایج سایر پژوهش ها مقایسه انجام می شود. در رابطه با سوال اول پژوهش یافته ها نشان داد که در این سال ها مولفین برتر حوزه ترجمه ماشینی به این شرح است. وی با ۷۵ اثر، سومیتا با ۷۲ و نی با ۵۸ اثر رتبه های اول تا سوم را به خود اختصاص داده اند. نکته جالب در این میان وجود تعداد زیادی محقق با خاستگاه جنوب شرق آسیاست که نهادینه شدن پژوهش در این حوزه را در جغرافیای آن منطقه نشان می دهد. اگرچه بسیاری از این محققین هم اکنون جذب سازمان های علمی اروپا و آمریکا شده اند و از آن وابستگی سازمانی استفاده می کنند ولی این نشان دهنده ماهیت سیال جریان علم است. این یافته ها با نتایج تحقیق ووس و ژایو (۲۰۰۵) مبنی بر حضور محققین مستقر در جوامع صنعتی در رتبه های بالای انتشارات ترجمه ماشینی همخوانی دارد.

سوال دوم به بررسی دانشگاه ها و کشورهای برتر در انتشارات حوزه ترجمه ماشینی می پردازد. یافته ها نشان داد که دانشگاه هایی که در کشورهای پیشرو در حوزه علم و فناوری های نوین قرار دارند در صدر موسساتی هستند که عمده تحقیقات این حوزه را منتشر می کنند. دانشگاه های دوبلین، کارنگی ملون و موسسه اطلاعات ملی ژاپن در صدر موسساتی هستند که بیشترین تولیدات این حوزه را عرضه می کنند.

جدول ۱ ده دانشگاه برتر را بر اساس تعداد مقالات نشان می دهد. این فهرست می تواند به خوانندگان علاقه مند کمک کند تا دانشگاه های فعال در زمینه تحقیقات ترجمه ماشینی را برای همکاری های پژوهشی بالقوه شناسایی

کنند. نتایج نشان داد که اغلب دانشگاه های برتر این حوزه در میان ۱۰ دانشگاه برتر در آمریکا واقع شده اند. این نشان دهنده سرمایه گذاری قابل توجه در تحقیقات ترجمه ماشینی و تعهد دانشگاه های ایالات متحده به این حوزه تحقیقاتی مهم است. البته با توجه به اینکه حوزه ترجمه ماشینی فناوری فنانورانه و نوآورانه دارد، بی دلیل نیست که حضور پرتعداد دانشگاه های کشورهای صنعتی مشهود است. این یافته با یافته های گاسپاری و دیگران (۲۰۱۵) که با نگاهی به روند تولید علم در جهان اذعان داشت کشورهای مثل آمریکا، انگلیس، ژاپن و چین در صدر آن قرار دارند همخوانی داشته و بیانگر ماهیت فنی و بسیار پیشرفته حوزه ترجمه ماشینی است. موفقیت در راه اندازی یک نظام ترجمه ماشینی کارآمد مستلزم دانش و ابزاری است که عموماً در جوامعی که زبانشناسی رایانه ای در آن، از پیشرفت مناسبی برخوردار است وجود دارد. نمونه بارز آن نظام ترجمه ماشینی گوگل است که معتبرترین نظام در این حوزه است.

در یافته های سوال سه که به بررسی توزیع موضوعی انتشارات ترجمه ماشینی می پردازد، نیز مشخص شد که رشته علوم کامپیوتر به عنوان یدر حوزه ترجمه ماشینی حرف اول را می زند. پیشرفت علوم کامپیوتری در کشورهای صنعتی بار دیگر یافته های این پژوهش را تایید کرد و با یافته های پژوهش هاچینز (۲۰۱۵) که علوم رایانه ای و علوم انسانی را پرتولیدترین حوزه های ترجمه ماشینی یافت همخوانی دارد. حوزه های علوم انسانی و هنر و علوم اجتماعی در رتبه های بعدی هستند و مشارکت بالایی در تحقیقات ترجمه ماشینی داشته اند. از آنجا که ترجمه ماشینی به دانش زیاد در حوزه علوم کامپیوتری نیاز دارد، بدیهی است که این حوزه در راس سایر موضوعات قرار گیرد. ضمن آنکه چون ترجمه خاستگاهی در علوم انسانی دارد رتبه آن نیز مورد قبول واقع می شود. نتایج بررسی موضوعی انتشارات این حوزه نشان داد که این حوزه ماهیتی بین رشته ای دارد و محققین با وابستگی های موضوعی متنوعی در این حوزه نقش ایفا می کنند. چنانکه بیشترین انتشارات ترجمه ماشینی خاستگاه علوم رایانه ای داشته و هنر و علوم انسانی، علوم اجتماعی، ریاضیات و مهندسی نیز در رتبه های بعدی هستند.

سوال چهارم به بررسی دیداری سازی این حوزه می پردازد. نقشه های علمی در خصوص ساختار و تکامل رشته ترجمه ماشینی اطلاعات ارزشمندی در اختیار می گذارند. شبکه حاصل از محاسبه هم رخدادی واژه ها برای ایجاد بینش در خصوص ساختار علمی و نحوه تکامل این حوزه دیداری سازی شد. شکل ۱ خوشه بندی موضوعی که ساختار کلی تحقیقات در حوزه

ترجمه ماشینی را نشان داد. در این گراف، برای نشان دادن متراکم ترین حوزه های موضوعی از طیف رنگ قرمز، زرد، سبز و آبی استفاده شده است که به ترتیب قرمز متراکم ترین و آبی کم تراکم ترین حوزه های موضوعی را نشان می دهد. خوشه موضوعی با نام ترجمه ماشینی آماری با بیشترین تناوب و تراکم نشان داد که رویکرد آماری در طراحی ماشین های ترجمه در حال حاضر رویکرد اصلی در تحقیقات این حوزه است. اگر چه وجود واژه هایی چون نحو و نظم واژگان و درخت واژگان در این خوشه نشان دهنده یکپارچگی این تحقیقات با اصول زبانشناسی رایانشی بود. خوشه موضوعی دیگر خوشه بازیابی اطلاعات است. وجود این خوشه در این شکل بیهوده نیست زیرا بازیابی اطلاعات بویژه در مبحث بازیابی بین زبانی وابستگی مستقیم با ترجمه ماشینی دارد. وجود اصطلاحاتی نظیر پرس و جو، تاپیک و سیستم پرسش و پاسخ موید شکل گیری هوشمندانه این خوشه و موفقیت رویکرد استفاده شده در این پژوهش است. در گراف مورد بحث اصطلاحاتی که هم رخدادی بالایی دارند در فاصله نزدیک تری نسبت به هم قرار می گیرند در حالی که واژه هایی که در نقشه از یکدیگر دورتر قرار گرفته اند، چندان قرابت موضوعی با هم نداشته اند.

خوشه ترجمه خودکار با دارا بودن اصطلاحاتی نظیر کاربر و ارتباطات، نشان دهنده اهمیت ترجمه ماشینی در تعاملات انسانی و ارتباطات بین المللی است. در واقع این خوشه وجه علوم انسانی و جامعه شناسی را در تحقیقات این حوزه نمایندگی می کند. خوشه طراحی و ارزیابی ترجمه ماشینی، دیگر خوشه موضوعی بود که میان رشته است و نیاز به تحقیقات بیشتر در آن حس می شود. بطور کلی این خوشه ها ساختار علمی حوزه ترجمه ماشینی را نشان می دهد. در نمای تراکمی واژه های موضوعی ترجمه ماشینی، حوزه تحقیقات ترجمه ماشینی آماری با رنگ قرمز، داغ ترین حوزه پژوهشی این حوزه است. تعاملات کاربر و ترجمه ماشینی و نقش آن در ارتباطات با رنگ زرد دومین حوزه مورد علاقه برای پژوهش بشمار می رود. پس از آن سایر حوزه هایی که با رنگ سبز مشخص شده اند از اهمیت تقریباً یکسانی برخوردارند. موضوعاتی نیز در حاشیه آبی قرار گرفته اند که کمتر از سایر حوزه های اشاره شده مورد توجه قرار گرفته است. از دیدگاه اشباع بودن حوزه های موضوعی باید گفت موضوعاتی که در قسمت قرمز و نارنجی قرار گرفته اند دیگر از جهت کثرت انتشارات به حد اشباع رسیده اند. این یافته ها با تحقیقات ووس و ژایو (۲۰۰۵) در بسیاری جهات از جمله موضوعات داغ پژوهشی همخوانی دارد.

## ۵-۳. نتیجه‌گیری کلی

در این پژوهش ساختار علمی و نحوه تکامل رشته ترجمه ماشینی مورد بررسی قرار گرفت. این پژوهش اولین مطالعه علم سنجی در زمینه ترجمه ماشینی در ایران است. لذا این مطالعه با ارائه بینش‌های تازه در مورد ساختار مفهومی تحقیق در حوزه ترجمه ماشینی از طریق تحلیل هم‌واژگانی مبتنی بر پایگاه اسکوپوس به عنوان یک پایگاه داده جامع به ادبیات علم سنجی ترجمه ماشینی کمک می‌کند. علاوه بر این، در حالی که این مطالعه تصویری جامع از وضعیت فعلی تحقیقات ترجمه ماشینی ارائه می‌دهد، نتایج می‌تواند راه را برای مطالعات آینده با روشن کردن شکاف‌های موجود در این زمینه و ارائه جهت برای تحقیقات آینده هموار کند. سازمان‌های درگیر در پژوهش این حوزه می‌توانند از این یافته‌ها برای ایجاد تغییرات در سیستم‌های ترجمه موجود استفاده کنند. در مقایسه با سایر زمینه‌های تحقیقاتی، آشکار شد که برخلاف بسیاری از رشته‌ها، مقالات کنفرانس در حوزه ترجمه ماشینی از مقالات نشریات بیشتر است. این نشان می‌دهد که این حوزه دارای پژوهشگرانی است که علاقمند به حضور در گردهمایی‌های علمی و اشتراک دانش از این طریق هستند. در این پژوهش عمدتاً مقالات انگلیسی زبان مورد بررسی قرار گرفت. مطمئناً این واقعیت وجود دارد که یک تحلیل علم‌سنجی باید مرور جامعی بر انتشارات حوزه داشته باشد. اما در دسترس نبودن مقالاتی که به زبان‌های محلی نگاشته می‌شوند این امکان را از پژوهشگر گرفت. مطمئناً نتایج این تحقیق خلاهای پژوهشی را در این حوزه مشخص کرده و از سوی دیگر نقاط داغ پژوهشی این حوزه را نشان می‌دهد. شناسایی محققین و دانشگاه‌های برتر این حوزه به پژوهشگران علاقمند کمک می‌کند که در جستجوی همکاری علمی، انتخاب‌های متعددی داشته باشند. یافته‌های پژوهش بویژه در بخش دیداری سازی، به پژوهشگران تازه کار این حوزه نمایی از ساختار علمی این حوزه را نشان می‌دهد. این پژوهشگران می‌توانند درک کنند که برای شروع پژوهش در این حوزه ضمن کسب دانش کلی در مورد ترجمه ماشینی، بایستی مطالعه خود را از کجا شروع کنند، کدام مراجع را بخوانند و کدام مجله را انتخاب کنند. آنها همچنین می‌توانند براساس تحلیل روندهای مطرح این حوزه، تصمیم بگیرند که کدام موضوع را انتخاب کنند. در نهایت می‌توان با اطمینان گفت که استفاده از روش تحلیل هم‌واژه برای شناسایی ساختار علمی حوزه ترجمه ماشینی مناسب بوده است. متأسفانه زبان فارسی در این تحقیقات در رتبه‌های برتر نیست. از مشکلات اصلی زبان فارسی در زمینه نرم افزارهای ترجمه ماشینی، ماهیت نحوی و معنایی آن است. زیرا ساختار متمایزی با زبان غالب علم یعنی زبان انگلیسی دارد. انگلیسی دارای جامعه پژوهشی بسیار بزرگتری بوده



و عمده نرم افزارهای پردازش زبان برای آن طراحی می شود. به منظور تسریع تحقیقات ترجمه ماشینی، زبان فارسی به یک سیاست بلندمدت با هدف اولویت بندی حوزه های تحقیق و توسعه در ترجمه ماشینی، شناسایی نقش بخش خصوصی در توسعه سیستم و شناسایی سازمان هایی که پتانسیل اصلی برای انجام تحقیقات ترجمه ماشینی دارند، نیاز دارد.

## ۴-۵ پیشنهادهایی برای پژوهش های آینده

بر اساس یافته های پژوهش، پیشنهادات زیر برای انجام پژوهش های آتی اعلام می گردد:

- این پژوهش با استفاده از پایگاه های استنادی وب علم و پایگاه استنادی علوم جهان اسلام نیز انجام شود و نتایج آن با یکدیگر تطبیق داده شود تا نمای کاملی از ساختار علمی حوزه ترجمه ماشینی به دست آید.
- دوره های زمانی جدیدتر نیز مورد بررسی قرار گیرد تا نحوه تغییر و تکامل سیر گرایش های پژوهشی مشخص شده در این تحقیق مشخص شده و با گرایش های در حال محو شدن احتمالی و گرایش های نوظهور نیز آشنایی حاصل شود.
- با مرور انتشارات کشورهای فارسی زبان و عربی زبان که رسم الخط تقریبا مشابهی دارند، به حوزه های تحقیقاتی ترجمه ماشینی این زبانها نیز ورود کرد تا چالش های پژوهش ترجمه ماشینی فارسی هم مشخص شود.
- تحلیل هم واژگانی تنها کلمات موجود در عنوان، چکیده و کلیدواژه های مقاله را تجزیه و تحلیل می کند و متون کامل را تجزیه و تحلیل نمی کند. بنابراین، تحقیقات آینده می تواند این مطالعه را عمیق تر مورد بررسی قرار دهد. به عنوان مثال، محققان می توانند دانشگاهها و موسساتی را که این مطالعات در آنها انجام شده است یا تفاوت آنها از نظر روش شناسی (کمی، کیفی یا ترکیبی)، جمع آوری داده ها و متغیرهای کلیدی مورد استفاده را بررسی کنند.
- پیشنهاد می شود به لحاظ فنی بودن حوزه ترجمه ماشینی، تحقیقی در زمینه ارتباط با صنعت و میزان تبدیل شدن دانش موجود به نرم افزار های ترجمه ماشینی انجام گردد.

## ۵-۵. پیشنهادهای کاربردی و برگرفته از نتایج پژوهش حاضر

- پیشنهاد می شود نتایج بدست آمده از پژوهش حاضر در اختیار پژوهشگاه‌ها و مراکز تحقیقاتی مرتبط با تحقیقات ترجمه ماشینی از جمله مراکز ایرانی مثل مترجم های الکترونیکی فارسی قرار گیرد.
- پیشنهاد می شود مجری طرح پژوهشی نتایج بدست آمده از پژوهش خود را در قالب سخنرانی و یا کارگاه برای سازمان‌ها، شرکت ها و نیز دانشگاه هایی که علاقمند به حوزه ترجمه ماشینی هستند ارائه نماید.
- پیشنهاد می شود موسسه استنادی علوم و پایش علم و فناوری به جهت داشتن گروه زبانشناسی رایانه ای، با استفاده از نتایج این تحقیق گام های آغازین طراحی نظام ترجمه ماشینی فارسی به سایر زبانها را بردارد.

# فهرست منابع

- Baker, M. and Maier, C. (2011) Ethics in Interpreter & Translator Training: Critical Perspectives. *The Interpreter and Translator Trainer*, 5, 1-14.  
<https://doi.org/10.1080/13556509.2011.10798809>
- Blei, D. M. & Lafferty, J. D. (2009). Topic models. In A. N. Srivastava & M. Sahami (Eds.), *Text Mining: Classification, Clustering, and Applications*, (pp. 71–94). London:
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of Machine Learning Research*, 3, 993-1022.
- Callon, M., Courtial, J.-P., Turner, W. A., & Bauin, S. (1983). From Translations to Problematic Networks: An Introduction to Co-word Analysis. *Social Science Information* 22(2), 191- 235.
- Norwood, NJ: Ablex. Danowski, J. A. (2009). Inferences from word networks in messages. In K. Krippendorff & M. A. Bock (Eds.), *The content analysis reader* (pp. 421-429). Los Angeles, etc.: Sage.
- de Miranda Santo, M., Coelho, G.M., dos Santos, D.M. and Fellows Filho, L., 2006. "Text mining as a valuable tool in foresight exercises: A study on nanotechnology". *Technological Forecasting and Social Change*, 73(8), pp.1013-1027.
- Diesner, J., Aleyasen, A., Chin, C., Mishra, S., Soltani, K., Tao, L. (2015). ConText: Network Construction from Texts [Software]. Available from <http://context.lis.illinois.edu/>
- Feldman, R., Fresko, M., Hirsh, H., Aumann, Y., Liphstat, O., Schler, Y. and Rajman, M., 1998. "Knowledge Management: A Text Mining Approach". In *PAKM* (Vol. 98, p. 9).

Garfield, E. 2001. "From computational linguistics to algorithmic historiography". In *Symposium in Honor of Casimir Borkowski at the University of Pittsburgh School of Information Sciences*.

Gaspari, F., Almaghout, H. and Doherty, S. (2015) A Survey of Machine Translation Competences: Insights for Translation Technology Educators and Practitioners. *Perspectives*, 23, 333-358. <https://doi.org/10.1080/0907676X.2014.979842>

gholi, f., 2014. "A survey of two decades of Iranian linguistics conference: a scientometric analysis". *Zabzn shenakht*.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, 520, 429-431.

Holzinger, A., Geierhofer, R., Mödritscher, F. and Tatzl, R., 2008." Semantic Information in Medical Information Systems: Utilization of Text Mining Techniques to Analyze Medical Diagnoses". *J. UCS*, 14(22), pp.3781-3795.

Hutchins, J., 2005. "The history of machine translation in a nutshell". Retrieved December, 20:2009, 2005.

Jacobi, C., van Atteveldt, W., & Welbers, K. (2015). Quantitative analysis of large amounts of journalistic texts using topic modeling. *Digital Journalism*, doi: 10.1080/21670811.2015.1093271

Krallinger, M., Valencia, A. and Hirschman, L., 2008." Linking genes to literature: text mining, information extraction, and retrieval applications for biology". *Genome biology*, 9(2), p.S8.

Moed, H.F., Glänzel, W. and Schmoch, U., 2004. "Editors' introduction". In *Handbook of quantitative science and technology research* (pp. 1-15). Springer, Dordrecht.

Rip, a., & Courtial, J.-P., 1984." Co-word maps of biotechnology: An example of cognitive scientometrics". *Scientometrics*, 6(6), 381-400.

Pym, A. (2014) Translation Skill-Sets in a Machine-Translation Age. *Meta*, 58, 487-503. <https://doi.org/10.7202/1025047ar>

Van Driel, M.A., Bruggeman, J., Vriend, G., Brunner, H.G. and Leunissen, J.A., 2006. "A text-mining analysis of the human phenome". *European journal of human*

*genetics*, 14(5), p.535.

VOS. and Zhao, X., 2005. "Some Steps Towards a Scientometric Analysis of Publications in Machine Translation". In *Artificial Intelligence and Applications* (pp. 651-655).



**Regional Information Center for Science and Technology**

**Final Report of the Research Project**

Title

**Study of the scientific structure and evolution of the field of machine translation in the years 2006 to 2015 based on Scopus database**

Director

**Hamid Alizadeh**

**January 2023**