

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

کلیه امتیازهای این پایان نامه به دانشگاه بوعلی سینا تعلق دارد. در صورت استفاده از تمام یا بخشی از مطالب این پایان نامه در مجلات، کنفرانس‌ها و یا سخنرانی‌ها، باید نام دانشگاه بوعلی سینا یا استاد راهنمای پایان نامه و نام دانشجو با ذکر مأخذ و ضمن کسب مجوز کتبی از دفتر تحصیلات تکمیلی دانشگاه ثبت شود. در غیر این صورت مورد پیگرد قانونی قرار خواهد گرفت. درج آدرس‌های ذیل در کلیه مقالات خارجی و داخلی مستخرج از تمام یا بخشی از مطالب این پایان نامه در مجلات، کنفرانس‌ها و یا سخنرانی‌ها الزامی می‌باشد.

....., Bu-Ali Sina University, Hamedan, Iran.

مقالات خارجی

..... گروه ..... دانشکده ..... دانشگاه بوعلی سینا، همدان.

مقالات داخلی



دانشگاه شاهرود

دانشکده مهندسی  
گروه آموزشی مهندسی کامپیوتر

پایان نامه برای دریافت درجه کارشناسی ارشد در رشته مهندسی کامپیوتر گرایش هوش مصنوعی

**عنوان:**

**ساخت وردنت فارسی در حوزه‌ی فاوا  
به کمک وردنت‌های دیگر**

استاد راهنما:

**دکتر محمد نصیری**

استاد مشاور:

**دکتر محرم منصوری‌زاده**

نگارش:

**اکرم احمدی طامه**

پاس بی کران پروردگار می تازد که هستی مان بخشد و به طریق علم و دانش رهنمون شود به بهنشین رحروان علم و دانش مستخرمان نمود و خوشه چینی از علم و معرفت را روزی مان ساخت.

## تقدیم به پدر بزرگوار و مادر مهربانم

ای پدر از تو هر چه می گویم باز هم کم می آورم، خورشیدی شدی و از روشنائی ات جان گرفتم و در ناامیدی مانازم را کشیدی و لبریزم کردی از شوق، اکنون حاصل دستان خسته ات رزم فزونی شد به خودم تبریک می گویم که تو را دارم.....

و تو ای مادر، ای شوق زیبای نفس کشیدن، ای روح مهربان هستی ام، تو رنگ شادی هایم شدی و لحظه ها را با تمام وجود از من دور کردی و عمری محنتی را به جان خریدی تا اکنون توانستی طعم خوش پیروزی را به من بچشانی.

## شکر از همسر عزیزم

به پاس قدردانی از قلبی آکنده از عشق و معرفت که محیطی سرشار از سلامت و امنیت و آرامش و آسایش برای من فراهم آورده است.

## تشکر و قدردانی

این پایان نامه با حمایت مالی موسسه تحقیقات ارتباطات و فناوری اطلاعات طی قرارداد شماره ۵۰۰/۶۶۱۴/ت مورخ ۹۰/۰۴/۲۸ انجام گرفته و بدینوسیله از این موسسه کمال تشکر و قدردانی را دارم.

بسی شایسته است از استادان فرهیخته و فرزانه جناب آقای دکتر محمد نصیری و جناب آقای دکتر محرم منصوری زاده که با کرامتی چون خورشید، سرزمین دل را روشنی بخشیدند و گلشن سرای علم و دانش را با راهنمایی‌های کار ساز و سازنده بارور ساختند، تقدیر و تشکر نمایم.



دانشگاه بوعلی سینا

مشخصات رساله/پایان نامه تحصیلی

عنوان:

ساخت وردنت فارسی در حوزه‌ی فاوا به کمک وردنت‌های دیگر

نام نویسنده: اکرم احمدی طامه

نام استاد راهنما: محمد نصیری

نام استاد مشاور: محرم منصوری‌زاده

دانشکده: مهندسی

گروه آموزشی: مهندسی کامپیوتر

رشته تحصیلی: مهندسی کامپیوتر

گرایش تحصیلی: هوش مصنوعی

مقطع تحصیلی: کارشناسی ارشد

تاریخ تصویب پروپوزال: ۱۳۹۰/۰۸/۳۰

تاریخ دفاع: ۱۳۹۱/۱۱/۲۳

تعداد صفحات: ۱۱۶

چکیده:

وردنت یک پایگاه داده‌ی لغوی بزرگ است که برای ابهام زدایی کلمات، بازیابی اطلاعات و ترجمه‌ی متون استفاده می‌شود. این پایگاه داده‌ی لغوی شامل اسم، فعل، صفت و قید است که کلمات هم‌معنی را در مجموعه‌هایی به نام ترادف دسته‌بندی می‌کند. در این پژوهش روش‌های خودکار را برای استخراج دادگان حوزه‌ی فناوری اطلاعات و ارتباطات (فاوا) از وردنت عمومی انگلیسی پرینستون پیشنهاد می‌کنیم و در مرحله‌ی بعد آن را به منظور ساخت وردنت فارسی نگاشت می‌دهیم. مزیت روش خودکار کاهش دخالت عامل انسانی و در نتیجه سرعت بخشی به فرایند توسعه وردنت فاوا می‌باشد. به همین منظور، چندین روش استخراج خودکار بر اساس تعریف ترادف، روابط معنایی بین ترادف‌های وردنت، شباهت بین تعریف ترادف‌ها و تعاریف لغت نامه‌های تخصصی موجود و بالاخره ترادف‌های برادر ارائه نموده‌ایم. نتایج پیاده‌سازی و ارزیابی این روش‌ها بر روی وردنت پرینستون نشان می‌دهد که استخراج دادگان فاوا دقت قابل قبولی به همراه داشته است.

واژه‌های کلیدی: وردنت، ترادف، ساخت خودکار، فاوا

## فهرست مطالب

پیشگفتار .....	۱
فصل اول: معرفی ساختار وردنت و حوزه ی فاوا .....	۷
۱.۱ مقدمه .....	۹
۱.۲ روشهای ساخت وردنت .....	۱۰
۱.۳ کاربردهای وردنت .....	۱۰
۱.۴ روابط معنایی وردنت ۳ انگلیسی .....	۱۱
۱.۴.۱ روابط عام-خاص .....	۱۱
۱.۴.۲ روابط کل-جزء .....	۱۲
۱.۴.۳ رابطه ی استلزام .....	۱۳
۱.۴.۴ رابطه ی علیت .....	۱۴
۱.۴.۵ رابطه ی شباهت .....	۱۴
۱.۴.۶ رابطه ی همچنین .....	۱۴
۱.۴.۷ رابطه ی خصلت .....	۱۴
۱.۴.۸ گروه بندی افعال .....	۱۵
۱.۴.۹ دامنه و عضو دامنه .....	۱۵
۱.۵ بررسی و نقد روابط معنایی در وردنت .....	۱۶
۱.۶ معرفی جداول پایگاه داده ی وردنت ۳ .....	۱۸
۱.۷ حوزه ی فناوری اطلاعات و ارتباطات .....	۲۰
۱.۸ جمع بندی .....	۲۰
فصل دوم: کارهای مرتبط .....	۲۱
۲.۱ مقدمه .....	۲۳
۲.۲ وردنت های عمومی .....	۲۳
۲.۲.۱ مروری بر وردنت ۳ .....	۲۳
۲.۲.۲ مروری بر وردنت فرانسوی .....	۲۷
۲.۲.۳ مروری بر وردنت دوزبانه ی چینی-انگلیسی .....	۲۹
۲.۲.۴ مروری بر وردنت ژاپنی .....	۳۲
۲.۲.۵ مروری بر وردنت روسیه ای (RussNet) .....	۳۴
۲.۲.۶ مروری بر وردنت چندگانه (MultiWordnet) .....	۳۷
۲.۲.۷ مروری بر وردنت کره ای .....	۳۸
۲.۲.۸ مروری بر وردنت استانفورد .....	۴۳
۲.۲.۹ مروری بر وردنت سوئدی و فنلاندی .....	۴۶
۲.۲.۱۰ مروری بر وردنت عربی .....	۴۷
۲.۲.۱۱ مروری بر فارسنت .....	۴۹
۲.۲.۱۲ یورونت وبالکانت .....	۵۱
۲.۳ وردنت های تخصصی .....	۵۷
۲.۳.۱ وردنت آلمانی (BIOWORDNET) .....	۵۸
۲.۳.۲ گسترش وردنت یونانی به حوزه ی روانشناسی و کامپیوتر .....	۵۸

۵۸	.....	۲.۴ جمع بندی
۶۱	.....	فصل سوم: روش‌های پیشنهادی
۶۳	.....	۳.۱ مقدمه
۶۵	.....	۳.۲ روش پیشنهادی اول: روش‌های مبتنی بر تعریف
۶۶	.....	۳.۲.۱ معیار شباهت کمینه
۶۷	.....	۳.۲.۲ معیار تطابق شباهت
۶۹	.....	۳.۳ روش پیشنهادی دوم: مبتنی بر روابط معنایی
۸۱	.....	۳.۴ روش پیشنهادی سوم: معیار LCS
۸۳	.....	۳.۵ روش پیشنهادی چهارم: معیار S3M
۸۵	.....	۳.۶ روش پیشنهادی پنجم: روش مبتنی بر رابطه‌ی معنایی Coordinate Term
۸۶	.....	۳.۷ جمع بندی
۸۷	.....	فصل چهارم: پیاده‌سازی و ارزیابی
۸۹	.....	۴.۱ مقدمه
۸۹	.....	۴.۲ ارزیابی روش مبتنی بر تعریف روی وردنت پرینستون
۸۹	.....	۴.۲.۱ شباهت کمینه
۹۰	.....	۴.۲.۲ ارزیابی روش تطابق شباهت
۹۱	.....	۴.۳ ارزیابی روش مبتنی بر روابط معنایی روی وردنت پرینستون
۹۲	.....	۴.۴ ارزیابی روش مبتنی بر معیار LCS
۹۲	.....	۴.۵ ارزیابی مبتنی بر معیار شباهت S3M
۹۳	.....	۴.۶ ارزیابی روش مبتنی بر رابطه‌ی Coordinate Term
۹۴	.....	۴.۷ ارزیابی کلی
۹۵	.....	۴.۸ جمع بندی
۹۷	.....	فصل پنجم: نتیجه‌گیری و سوی کارهای آتی
۹۹	.....	۵.۱ نتیجه‌گیری و سوی کارهای آتی



## فهرست جدول‌ها

- جدول ۱-۱ - کلیه روابط معنایی وردنت پرینستون به همراه نمونه مثال ..... ۱۷
- جدول ۱-۲ - بسامد واژگان مقوله‌های مختلف دستوری در وردنت پرینستون ..... ۲۶
- جدول ۲-۲ - تعداد ترادف‌های بدست آمده از منابع مختلف در وردنت ولف ..... ۲۸
- جدول ۲-۳ - ارزیابی اتوماتیک وردنت ولف ..... ۲۹
- جدول ۳-۱ - لیست مرجع واژگانی فاوا ..... ۶۶
- جدول ۳-۲ - لیست واژگانی با درجه‌ی تکنیکی ۱ ..... ۷۳
- جدول ۳-۳ - لیست واژگانی با درجه‌ی تکنیکی ۲ ..... ۷۶
- جدول ۳-۴ - لیست واژگانی با درجه‌ی تکنیکی ۳ ..... ۷۸
- جدول ۳-۵ - لیست واژگانی با درجه‌ی تکنیکی ۴ ..... ۷۹
- جدول ۳-۶ - لیست واژگانی با درجه‌ی تکنیکی ۵ ..... ۸۱
- جدول ۴-۱ - نتایج ارزیابی روش مینیمم شباهت ..... ۹۰
- جدول ۴-۲ - نتایج ارزیابی روش تطابق شباهت ..... ۹۰
- جدول ۴-۳ - ارزیابی روش مبتنی بر روابط معنایی ..... ۹۱
- جدول ۴-۴ - ارزیابی روش مبتنی بر معیار LCS ..... ۹۲
- جدول ۴-۵ - ارزیابی روش مبتنی بر معیار S<sup>3</sup>M ..... ۹۲
- جدول ۴-۶ - ارزیابی روش مبتنی بر رابطه‌ی Coordinate Term ..... ۹۳
- جدول ۴-۷ - ارزیابی همه‌ی روش‌های پیشنهادی ..... ۹۴

## فهرست شکل‌ها

- شکل ۱-۱ -۱- رابطه‌ی Hypernym/Hyponym ..... ۱۱
- شکل ۱-۲ -۲- رابطه‌ی Instance Hypernym/Instance Hyponym ..... ۱۲
- شکل ۱-۳ -۳- رابطه‌ی Part holonym/Part Meronym ..... ۱۲
- شکل ۱-۴ -۴- رابطه‌ی Member holonym/Member Meronym ..... ۱۳
- شکل ۱-۵ -۵- رابطه‌ی Domain Category/Domain Member Category ..... ۱۵
- شکل ۲-۱ -۱- نمودار بسامد واژگان مقوله‌های مختلف دستوری در وردنت ۳ انگلیسی ..... ۲۶
- شکل ۲-۲ -۲- رابطه‌ی IS-A ..... ۴۰
- شکل ۲-۳ -۳- مرحله‌ی آموزش ..... ۴۲
- شکل ۲-۴ -۴- مرحله‌ی نگاشت ..... ۴۳
- شکل ۳-۱ -۱- فلوچارت روش مینیمم شباهت ..... ۶۷
- شکل ۳-۲ -۲- فلوچارت روش مبتنی بر معیار تطابق شباهت ..... ۶۸
- شکل ۳-۳ -۳- ساختار استخراج ترادف‌های حوزه‌ی فاوا با روش مبتنی بر روابط معنایی ..... ۷۲
- شکل ۳-۴ -۴- نمونه مثال برای استخراج دنباله‌ی کلمات ..... ۷۳
- شکل ۳-۵ -۵- نحوه‌ی استخراج ترادف با استفاده از نمونه مثال ..... ۷۶
- شکل ۳-۶ -۶- فلوچارت روش مبتنی بر LCS ..... ۸۲
- شکل ۳-۷ -۷- فلوچارت روش مبتنی بر روش  $S^3M$  ..... ۸۴
- شکل ۳-۸ -۸- نمایش ساختار مبتنی بر رابطه‌ی Coordinate Term ..... ۸۵
- شکل ۳-۹ -۹- نمونه مثال برای استخراج ترادف از طریق رابطه‌ی Coordinate Term ..... ۸۶

پسکھار



شبکه‌ی معنایی<sup>۱</sup> یکی از مشهورترین فرم‌های نمایش اطلاعات ساخت‌یافته است. این شبکه همانند سایر شبکه‌ها از گره‌ها و پیوندها ساخته شده است که گره‌ها همان مفاهیم و پیوندها نیز ارتباط بین این مفاهیم هستند. یک شکل خاص از شبکه‌های سنتی معنایی، کاری است که جورج میلر<sup>۲</sup> و همکارانش در دانشگاه پرینستون انجام داده‌اند [1]. آنها مفهوم شبکه‌ی معنایی لغوی را گسترش دادند و وردنت<sup>۳</sup> را ساختند. درواقع وردنت یک پایگاه داده لغتی بزرگ است که شامل اسم، فعل، صفت و قید و همچنین کلمات مترادف و متضاد می‌باشد. میلر و تیمش وردنت را که در ابتدا هدف آن تدوین یک فرهنگ لغت معنایی بود، گسترش دادند. آنها مجموعه‌ای از کلمه‌های هم معنی را در زبان انگلیسی گردآوری کردند که این مجموعه را یک ترداف<sup>۴</sup> نامیدند. کار آنها یک شبکه معنایی از مفاهیم است که هر مفهوم توسط رابطه بین تعدادی از ترداف‌ها در زبان انگلیسی بیان می‌شدند.

این پروژه از سال ۱۹۸۵ آغاز شده و تاکنون ادامه دارد و یافته‌های آن، راه را تا حدود زیادی برای تهیه‌ی وردنت‌های زبان‌های دیگر هموار ساخته است. در حال حاضر وردنت برای زبان‌های فرانسه، آلمانی، ایتالیایی، بلغاری، چینی، کره‌ای، یونانی، ترکی، عبری و حتی سانسکریت ارائه شده و همچنین انجمنی تحت عنوان انجمن جهانی وردنت تشکیل شده است. هدف این انجمن به دست آوردن استانداردهایی برای تهیه وردنت برای زبان‌های مختلف است [2].

در چند سال اخیر در داخل و خارج از کشور فعالیت‌هایی برای تهیه‌ی وردنت فارسی آغاز شده است. با توجه به کاربردهایی که امروزه برای وردنت در زمینه‌ی بازیابی اطلاعات و حوزه‌های مختلف پردازش معنایی تعریف شده است، علاوه بر زبان‌شناسان، توجه بسیاری از متخصصان علوم کامپیوتر نیز به تهیه‌ی وردنت فارسی معطوف شده است. نخستین نسخه از وردنت فارسی توسط دانشگاه شهید بهشتی و با حمایت مرکز تحقیقات مخابرات ایران ساخته شده است [3]. این محصول

---

<sup>1</sup>Semantic Network

<sup>2</sup>George Miller

<sup>3</sup>WordNet

<sup>4</sup>Synset

در بسیاری کاربردهای پردازش زبان فارسی از جمله ترجمه ماشینی، خلاصه سازی اخبار، جستجو و بازیابی اطلاعات، کشف هرزنامه ها، تحلیل اطلاعات متون و رمزنگاری معنایی نقش کلیدی بازی می-کند.

وردنت‌های تخصصی نیز همچون وردنت یونانی در حوزه‌ی کامپیوتر و روانشناسی ساخته شده است. با توجه به رشد و توسعه فناوری اطلاعات و ارتباطات (فاوا) و ورود آن به حوزه‌ی عمومی، واژگان فاوا به تدریج به واژگان عمومی افزوده می‌شوند. از این رو ایده‌ی این پایان‌نامه مبنی بر جستجو برای واژگان مرتبط با حوزه‌ی فاوا در نسخه ۳ وردنت شکل گرفت.

به دلیل کمبود متون معتبر فارسی در حوزه فاوا، ابتدا وردنت انگلیسی ساخته شده و سپس به زبان فارسی ترجمه می‌شود. در این تحقیق با استفاده از وردنت پرینستون ترادف‌های مربوط به حوزه‌ی فاوا استخراج می‌شوند. برای این کار، ابتدا با روش‌های دسته بندی متن ترادف‌ها را به دو دسته‌ی فاوا و غیرفاوا تقسیم کردیم. بعد از استخراج، ترادف‌ها و روابط معنایی مربوطه به وردنت حوزه‌ی فاوا نگاشت داده می‌شود و وردنت فارسی حوزه‌ی فاوا از طریق ترجمه‌ی دادگان فاوای انگلیسی ایجاد می‌شود. روش‌هایی برای دسته‌بندی ترادف‌های وردنت پیشنهاد شده است که در ادامه توضیح داده می‌شود.

یک روش پیشنهادی جهت جمع‌آوری دادگان فاوا، روش استخراج خودکار از وردنت پرینستون است که چندین روش را برای آن پیاده‌سازی و ارزیابی کرده‌ایم. مزیت روش خودکار کاهش دخالت عامل انسانی و در نتیجه سرعت بخشی به فرایند توسعه وردنت فاوا است. یکی از روش‌های پیشنهادی پیاده‌سازی شده مبتنی بر تعاریف ترادف‌های وردنت و با استفاده از مجموعه واژگان فاوا است. در این روش با حذف و اضافه کردن کلماتی به لیست واژگان مرجع فاوا سعی کردیم میزان دقت را افزایش دهیم هر چند این افزایش دقت منجر به کاهش میزان دادگان فاوای استخراج شده می‌شود.

روش پیشنهادی دوم بر اساس روابط معنایی وردنت و با استفاده از مجموعه واژگانی با درجه تکنیکی<sup>۵</sup> متفاوت پیاده‌سازی شده است. درجه‌ی تکنیکی به میزان رایج بودن کلمه در حوزه فاوا مربوط می‌شود. در آزمایشات مربوط به این روش ما نشان می‌دهیم که هر چه درجه‌ی تکنیکی کلمات افزایش می‌یابد به علت کاهش میزان عمومیت واژه‌ها میزان دقت افزایش می‌یابد ولی از طرفی به علت افزایش درجه‌ی تکنیکی و کاهش کاربرد آن‌ها در وردنت پرینستون تعداد کمتری مترادف استخراج می‌شود.

روش سوم و چهارم بر اساس معیار شباهت‌های LCS و S<sup>3</sup>M بین تعاریف مترادف‌ها و تعاریف لغت‌نامه‌ی مایکروسافت است. در این روش‌ها سعی بر این است که مترادف‌هایی از وردنت که دارای میزان شباهتی بیشتر از یک حد آستانه هستند به عنوان مترادف‌های کاندید حوزه‌ی فاوا استخراج شوند. در این دو روش یک دید هم‌تراز و کلی بین مترادف‌های وردنت وجود دارد یعنی هیچ مجموعه از کلمات تخصصی فاوا به عنوان محدودیت بر تعریف مترادف‌ها یا کلمات مترادف‌ها اعمال نمی‌شود. در ارزیابی مربوط به این روش‌ها دیده می‌شود که مترادف‌هایی که مختص این روش‌ها بوده و توسط روش‌های دیگر استخراج نشده‌اند دارای کلمات تخصصی فاوا در تعریف مترادف یا در کلمات مترادف نیستند.

روش پیشنهادی آخر بر اساس رابطه‌ی Coordinate در وردنت تعریف شده است. این رابطه به صورت فرزندان یک گره پدر تعریف می‌شود. یعنی دو گره فرزند از یک گره پدر با هم رابطه‌ی Coordinate دارند. این رابطه را بر روی مترادف‌های حوزه‌ی فاوا که در روش‌های پیشین استخراج شده بود، اعمال کردیم تا بتوانیم مترادف‌های بردار با مترادف‌های فاوای اولیه را استخراج کنیم.

در ادامه ابتدا در فصل اول به بررسی ساختار و روابط معنایی استفاده شده در وردنت پرینستون، روش‌های ساخت وردنت و تعریف حوزه‌ی فاوا پرداخته می‌شود. در فصل دوم به مرور

---

<sup>5</sup>TechFactor

---

کارهای انجام شده و وردنت‌های ساخته شده تا کنون می‌پردازیم. در این فصل وردنت‌های تخصصی و عمومی را به صورت جداگانه مورد بحث قرار می‌گیرند. در فصل سوم روش‌های پیشنهادی برای استخراج خودکار دادگان حوزه‌ی فاوا شرح داده می‌شود. در این فصل پنج روش را با شرح کامل ساختار الگوریتم‌هایشان و به همراه بهبودهایی که در بعضی از آن روش‌ها حاصل شده است، ارائه می‌دهیم. در فصل چهارم نتایج حاصل از این روش‌ها را ارزیابی نموده و به بررسی نتایج و دقت آنها می‌پردازیم. در نهایت در فصل پنجم به نتیجه‌گیری و سوی کارهای آینده می‌پردازیم.



# ۱ فصل اول

معرفی ساختار وردنت و

حوزه می فاوا



## ۱.۱ مقدمه

وردنت یک پایگاه داده لغوی بزرگ از لغات است. این بانک اطلاعاتی اسم‌ها، فعل‌ها، صفت‌ها و قیدها را به مجموعه‌ای از لغات مترادف<sup>۶</sup> دسته‌بندی می‌نماید که هر مجموعه ترادف یک مفهوم مجزا را بیان می‌کند. مجموعه ترادف‌ها با استفاده از روابط معنایی-مفهومی<sup>۷</sup> و ارتباط‌های لغوی<sup>۸</sup> به یکدیگر پیوند داده شده‌اند. شبکه بدست آمده شبکه‌ای از لغات و مفاهیم معنایی مرتبط است که می‌تواند توسط مرورگرها پیمایش شود.

می‌توان گفت که وردنت مشابه یک لغت نامه است که لغات را بر اساس معانی‌شان دسته‌بندی می‌کند. با این حال چندین تفاوت مهم بین وردنت و دیگر لغت نامه‌ها وجود دارد. اول اینکه وردنت تنها شکل کلمات - رشته‌هایی از حروف - را پیوند نمی‌دهد، بلکه مفاهیم لغات را نیز مرتبط می‌سازد. در نتیجه لغاتی که در نزدیکی یکدیگر در شبکه یافت می‌شوند، قرابت معنایی نیز دارند. دومین تفاوت این است که وردنت روابط معنایی میان لغات را برچسب گذاری می‌کند، در حالی که دسته‌بندی‌های لغات در یک لغت نامه از هیچگونه الگوی مشخصی جز مشابهت معنایی پیروی نمی‌نماید.

ایجاد وردنت دو هدف دارد:

- ایجاد فرهنگ جامع لغات.
- تحلیل و کاوش متون به طور خودکار و با استفاده از الگوریتم‌های هوش مصنوعی.

از آنجایی که وردنت حوزه‌ی فاوا برای کاربران این حوزه در زمینه‌های پردازش زبان طبیعی و ترجمه‌ی ماشینی مفید و کاربردی است و از طرفی یک فرهنگ از لغات تخصصی را در اختیار آنها قرار می‌دهد لذا ساخت چنین وردنتی را بر اساس وردنت پرینستون برگزیدیم.

<sup>۶</sup>synonymy

<sup>۷</sup>conceptual-semantic

<sup>۸</sup>lexical

جهت ساخت وردنت فاوا بر اساس وردنت پریستون، نیاز به شناخت و بررسی ساختار، جداول و روابط معنایی وردنت انگلیسی نسخه‌ی ۳ است. لذا در ادامه‌ی فصل بعد از بررسی انواع روش ساخت و مزایای ساخت وردنت، روابط معنایی و جدول‌های وردنت انگلیسی نسخه‌ی ۳ تحلیل و بررسی می‌شود. در آخر فصل توضیح مختصری از حوزه‌ی فاوا بر اساس فرهنگ لغات، سایت‌های مرجع و کتابهای تخصصی ارائه می‌شود.

## ۱.۲ روش‌های ساخت وردنت

ساخت وردنت می‌تواند به سه صورت امکان پذیرد:

- دستی
- نیمه‌خودکار
- خودکار

در روش دستی با استفاده از دانش خود زبان به استخراج اطلاعات و ساخت وردنت پرداخته می‌شود. در روش خودکار از نگاشت وردنت‌های دیگر و مستقل از قوانین خود زبان به ساخت وردنت پرداخته می‌شود. در روش نیمه خودکار قسمتی از روند کار بر اساس نگاشت انجام می‌گیرد و در ادامه یک پالایش و ویرایش دستی روی کارهای انجام شده صورت می‌گیرد.

ساخت دستی وردنت که شامل بخش عظیم ایجاد روابط معنایی است، وقت‌گیر و مستعد خطاست و به دانش مربوط به آن زبان نیاز دارد ولی ساخت خودکار دخالت عامل انسانی را کاهش و سرعت ساخت را افزایش می‌دهد.

## ۱.۳ کاربردهای وردنت

حوزه‌های کاربردی وردنت عبارتند از :

- ابهام‌زدایی مفهوم کلمات