



Corpus-based Classification of Persian Homographs

Elham Alayiaboosar

PhD in General Linguistics; Assistant Professor;
Iranian Research Institute for Information Science and Technology
(IranDoc); Tehran, Iran Email: alayi@irandoc.ac.ir

Received: 18, Apr. 2020

Accepted: 24, Jan. 2021

Abstract: One of the big challenges in natural language processing is ambiguity. Homographs and homograph sense disambiguation is highly important in computational processing of texts. In languages with complex morphology, there exists lots of homographs, which are worth studying and classifying. In present study, in order to study Persian homographs extracted from the corpus, first words with more than one POS tag were extracted from an annotated corpus, 10978 words. Then, the frequency of each POS tag pertaining to every homograph was studied and another list of homographs was extracted from the first one, which include homographs with high frequency of the first tag (more than 20) and considerable frequency of the second tag (more than 10); the mentioned list include 1675 homographs. Morphological, phonological as well as semantic structures of homographs were studied based on which all homographs were classified into 11 categories. From 11 categories only homographs in one category were chosen based on semantic criteria and the rest were classified based on morphological as well as phonological criteria. The output of the present study includes a big list of homographs extracted from Persian text corpus, each of which are categorized in one or more than one category based on morphological and phonological characteristics of homographs. Such list and the related categorization could be used in word sense disambiguation systems.

Keywords: Ambiguity, Homograph, POS Tag, Morphological Characteristics

دسته‌بندی پیکره‌بنیاد هم‌نگاره‌های فارسی

الهام علایی ابوذر

دکتری زبان‌شناسی؛ استادیار؛
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)؛
تهران، ایران alayi@irandoc.ac.ir



مقاله برای اصلاح به مدت ۹ روز نزد پدیدآور بوده است.

پذیرش: ۱۴۰۱/۰۴/۰۱

دریافت: ۱۴۰۱/۰۱/۲۰

نثریه علمی | رتبه بین‌المللی
پژوهشگاه علوم و فناوری اطلاعات ایران
(ایرانداک)

شاپا (چاپی) ۸۲۲۳-۲۲۵۱

شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱

نمایه در SCOPUS، ISI، LISTA و

jipm.irandoc.ac.ir

دوره ۳۸ | شماره ۳ | صص ۸۲۵-۹۰۰

بهار ۱۴۰۲

<https://doi.org/jipm.38.3>



چکیده: ابهام، یکی از چالش‌های بزرگ در پردازش زبان طبیعی است. در پردازش رایانه‌ای متون، هم‌نگاره‌ها و چگونگی رفع ابهام از آن‌ها از اهمیت بالایی برخوردار است. در زبان‌هایی مانند زبان فارسی که ساخت‌واژه پیچیده‌ای دارند، هم‌نگاره‌های بسیاری ساخته می‌شوند. در این راستا شناخت هم‌نگاره‌ها و دسته‌بندی انواع آن‌ها بسیار مهم است. در پژوهش حاضر به‌منظور بررسی پیکره-بنیاد هم‌نگاره‌های فارسی، واژه‌هایی که بیش از یک برچسب اجزای واژگانی کلام داشتند، از پیکره متنی فارسی استخراج شدند که شامل ۱۰۹۷۸ واژه است. سپس، فراوانی برچسب‌های هر هم‌نگاره مورد بررسی قرار گرفت و فهرست دیگری استخراج شد که شامل هم‌نگاره‌هایی است که افزون بر فراوانی بالای برچسب اول آن‌ها (بیش از ۲۰)، فراوانی برچسب دوم آن‌ها نیز در پیکره متنی فارسی بیش از ۱۰ بوده است. این فهرست شامل ۱۶۷۵ هم‌نگاره است. ماهیت ساخت‌واژی، آوایی یا معنایی هم‌نگاره‌های استخراج‌شده در این مرحله مورد بررسی قرار گرفت و بر اساس آن، هم‌نگاره‌های استخراج‌شده در یازده دسته طبقه‌بندی شد که از میان آن‌ها، تنها هم‌نگاره‌های موجود در یک دسته بر اساس معیار معنایی طبقه‌بندی شده‌اند و بقیه بر اساس ویژگی‌های ساخت‌واژی و تفاوت‌های آوایی آن‌ها در دسته‌های گوناگون قرار گرفته‌اند. نتایج این پژوهش شامل فهرست گسترده‌ای از هم‌نگاره‌های فارسی استخراج‌شده از پیکره متنی فارسی است که هر کدام بر اساس معیارهای ساخت‌واژی و آوایی در یک یا بیش از یک دسته از مجموع یازده دسته قرار گرفته‌اند. این فهرست و دسته‌بندی مربوط به هم‌نگاره‌ها می‌تواند در سیستم‌های رفع ابهام معنایی از واژگان مورد استفاده قرار گیرد.

کلیدواژه‌ها: ابهام، هم‌نگاره، برچسب اجزای واژگانی کلام، ویژگی‌های ساخت‌واژی هم‌نگاره‌ها

۱. مقدمه

ابهام^۱ در زبان، یکی از چالش‌های بزرگ در پردازش زبان طبیعی^۲ است. بیشتر ابهام نوعاً در گونه نوشتاری زبان مد نظر است. ابهام در سطوح مختلف زبانی اتفاق می‌افتد (واژگان، جمله، پاراگراف و غیره). «ابهام» در فارسی برای معادل واژه ambiguity به کار می‌رود و در معناشناسی «گنگی» یا «گنگ بودن» برای vagueness مورد استفاده قرار می‌گیرد. «گاستاد» بر این باور است که دو واژه ambiguity و vagueness متفاوت هستند. وی vagueness را حالتی در نظر می‌گیرد که یک واژه یا عبارت، تنها یک معنا دارد که مرزهای آن به طور دقیق مشخص نشده است؛ در صورتی که ambiguity، حالتی است که یک واژه می‌تواند چندین معنا در بافت‌های گوناگون داشته باشد (Gaugstad 2004). در زبان‌شناسی رایانشی وقتی سخن از ابهام و رفع ابهام از معنای واژه‌ها^۳ و جمله‌هاست، منظور ambiguity است.

ابهام، انواع گوناگونی دارد که معروف‌ترین آن‌ها این موارد است: (۱) ابهام ساخت‌واژی^۴: این نوع ابهام به ساخت درونی واژه‌ها و چگونگی اتصال وندها یا بخش‌هایی از کلمات مرکب مربوط می‌شود. در برخی زبان‌ها ساخت درونی کلمات مرکب می‌تواند چنین ابهامی را ایجاد کند و شکل تقطیع واژه به تکواژهای تشکیل‌دهنده می‌تواند تعیین‌کننده معنای واژه باشد. به عنوان نمونه، کلمه مرکب computertaalkunde در زبان هلندی را به دو طریق می‌توان تجزیه کرد: computer+taalkunde، به معنای زبان‌شناسی رایانه‌ای و computertaal+kund، به معنای برنامه‌ریزی دانش زبانی (Gaugstad 2004). (۲) نوع دیگر ابهام که فراوانی آن در پیکره‌های زبانی قابل توجه است، ابهام نحوی^۵ است. نمونه بارز چنین ابهامی در زبان‌های گوناگون به طور معمول، هنگامی رخ می‌دهد که گروه‌های حرف اضافه‌ای به جمله اضافه می‌شوند. جمله معروف The man saw the girl with the telescope (مرد دختر را با تلسکوپ دید)، نمونه بارز ابهام نحوی است. (۳) از دیگر موارد ابهام می‌توان به ابهام معنایی واژگانی^۶ اشاره کرد و آن، زمانی رخ می‌دهد که یک صورت نوشتاری واحد به بیش از یک معنا دلالت کند؛ مانند واژه «شانه» در دو جمله «موهایم را شانه کردم» و «شانه‌ام درد گرفت» یا واژه «شیر» در سه

1. ambiguity

2. natural language processing (NLP)

3. word sense disambiguation (WSD)

4. morphological ambiguity

5. syntactic ambiguity

6. lexical semantic ambiguity

جمله «شیر سلطان جنگل است»، «شیر حاوی کلسیم است» و «شیر آب را ببندید». اینکه ما چگونه معنای یک واژه را درک می‌کنیم، بستگی دارد به عواملی چون بافت زبانی که آن واژه در آن به کار رفته است، دانش ما درباره جهان، و چگونگی کاربرد آن واژه در جامعه. واژه‌ها در طول زمان، دستخوش تغییرات معنایی، آوایی یا کاربردی می‌شوند. به عبارت دیگر، معنای واژه‌ها می‌تواند طی زمان تغییر کند یا در یک حوزه خاص یک معنا داشته باشد و طی زمان در حوزه‌ای دیگر و متفاوت دچار تغییرات معنایی شود. سه واژه تخصصی که در چنین شرایطی به کار می‌رود، هم‌نگاره^۱، هم‌آوا^۲، هم‌نام^۳ و چندمعنایی^۴ است. هم‌نگاره‌ها/هم‌نگاشت‌ها واژه‌هایی هستند که شکل نوشتاری یکسان، اما معنا، منشأ، مقوله واژگانی یا تلفظ متفاوت دارند (Merriam-Webster Dictionary) مانند «کرد» /kard/ و «کرد» /kord/، «مردم» /mardom/ و «مردم» /mordam/ یا «سیر» /sir/ و «سیر» /sejr/. هم‌آواها واژه‌هایی هستند که تلفظ یکسان اما معنا یا صورت نوشتاری متفاوت دارند (Merriam-Webster Dictionary)؛ مانند «خار» /xɑr/ و «خوار» /xɑr/. هم‌نام‌ها واژه‌هایی هستند که صورت نوشتاری و تلفظ یکسان، اما معنای متفاوتی دارند؛ مانند واژه «شیر» که قبلاً ذکر شد و چندمعنایی به حالتی اطلاق می‌شود که یک واژه بیش از یک معنا داشته باشد (Merriam-Webster Dictionary)؛ مانند واژه «قلب» /Galb/ در جمله‌ها و عبارت‌های «قلبش به درد آمد» و «قلب ماجرا».

در پردازش رایانه‌ای متون، هم‌نگاره‌ها و چگونگی رفع ابهام از آن‌ها از اهمیت بالایی برخوردار است و اساساً وقتی صحبت از رفع ابهام از معنای واژه‌هاست، به طور عمده، منظور رفع ابهام از هم‌نگاره‌هاست. در زبان‌هایی مانند زبان فارسی که ساخت واژه پیچیده‌ای دارند، هم‌نگاره‌های بسیاری ساخته می‌شوند. هم‌نگارگی در زبان فارسی ناشی از بازنمایی واجی و صرفی عناصر زبانی در خط فارسی است؛ به این صورت که رابطه‌های چند-به-چند و در برخی موارد، غیرنظام‌مند میان عناصر واجی و صرفی زبان فارسی و تظاهر نوشتاری آن‌ها در خط فارسی دیده می‌شود. اطلاق هم‌نگاره به چنین ساخت‌هایی تنها فارغ از بافت صورت می‌پذیرد، زیرا بافت نحوی باعث رفع ابهام از چنین هم‌نگاره‌هایی می‌شود. گویشوران باسواد فارسی در مواجهه با هم‌نگاره‌ها، به دلیل آشنایی با ساخت واژه، ساخت واجی و نحوی و هم‌آیندها^۵ در فارسی، قادر به رفع

1. homograph

2. homophone

3. homonym

4. polysemy

5. collocation

ابهام از هم‌نگاره‌ها هستند و در واقع، توانایی کاربرد صحیح هم‌نگاره‌ها را در بافت نحوی دارند. حال آنکه سامانه‌های پردازش متن اگر به این اطلاعات مجهز نباشند در پردازش متن، تخصیص برچسب صحیح به کلمات، ترجمه ماشینی و ... دچار خطاهای غیرقابل چشم‌پوشی می‌شوند.

در پژوهش حاضر سعی شده است دسته‌بندی پیکره-بنیاد از هم‌نگاره‌های فارسی ارائه گردد. امروزه، بسیاری از پژوهش‌های زبانی به کمک پیکره‌های زبانی انجام می‌شود. در واقع، زبان‌شناسی پیکره‌ای تأکیدی است بر کار کردن روی داده‌های واقعی زبان که از کاربرد طبیعی زبان در بافت به‌دست آمده و در یک پیکره جمع‌آوری شده است. در زبان‌شناسی پیکره فراوانی، پارامتری مهم برای کشف الگوهای متناوب و بالفعل است. در این راستا در پژوهش حاضر از پیکره متنی فارسی برای جست‌وجوی هم‌نگاره‌های فارسی استفاده شده است و سپس، دسته‌بندی هم‌نگاره‌ها با تأکید بر میزان فراوانی هر دسته انجام شده است. پیکره متنی زبان فارسی مجموعه‌ای از متون نوشتاری و گفتاری رسمی زبان فارسی است که از منابع واقعی همچون روزنامه‌ها، سایت‌ها و مستندات از قبل تایپ‌شده جمع‌آوری شده، تصحیح گردیده و برچسب خورده است. حجم این دادگان در حدود ۱۰۰ میلیون واژه است و از منابع مختلف تهیه گردیده و دارای تنوع بسیار زیادی است. ۱۰ میلیون واژه از این پیکره با استفاده از ۸۸۲ برچسب نحوی-معنایی به‌صورت دستی توسط دانشجویان رشته زبان‌شناسی برچسب‌دهی شده‌اند و هر پرونده بر حسب موضوع و منبع آن طبقه‌بندی شده است (Bijankhan et al. 2011). علت انتخاب این پیکره از میان انبوه پیکره‌های گوناگون تک‌زبان، دوزبان، عمومی و تخصصی موجود در زبان فارسی این است که پیکره متنی فارسی پیکره‌ای است عمومی که منبع مناسبی برای بررسی هم‌نگاره‌ها در بافت طبیعی زبان محسوب می‌شود. همچنین، برچسب‌های اجزای واژگانی کلام^۱ نیز این امکان را فراهم می‌آورد که بتوان واژه‌هایی را که بیش از یک برچسب دارند و به‌نوعی هم‌نگاره محسوب می‌شوند، از پیکره استخراج کرد و به دسته‌بندی آن‌ها پرداخت. در روش پژوهش چگونگی استفاده از پیکره توضیح داده می‌شود.

در این مقاله، پس از ارائه پیشینه پژوهش به چگونگی استخراج و دسته‌بندی هم‌نگاره‌ها از پیکره متنی فارسی پرداخته می‌شود و در انتها، بحث و نتیجه‌گیری ارائه می‌گردد.

1. parts of speech (POS)

۲. پیشینه پژوهش

در هر زبانی واژه‌های مبهمی وجود دارند که دارای معانی متفاوتی هستند. یافتن معنای درست واژه‌های دارای معانی متعدد (رفع ابهام)، مسئله‌ای مهم در پردازش زبان طبیعی محسوب می‌شود. انتخاب معنای درست واژه مبهم ممکن است برای انسان بسیار واضح و آسان باشد، اما تشخیص اینکه کدام یک از معانی موجود برای یک واژه در جمله‌ای خاص انتخاب شود، برای ماشین دشوار است. به‌منظور رفع ابهام از هم‌نگارها، ابتدا لازم است انواع هم‌نگارها و دسته‌بندی آن‌ها در زبان مورد بررسی مشخص شود. برخی پژوهشگران حوزه پردازش متن و بازیابی اطلاعات، بسته به نوع پژوهش خود، اقدام به دسته‌بندی هم‌نگارها در زبان مورد بررسی کرده‌اند. به‌عنوان نمونه، «یاروسکی» در پژوهش خود (رفع ابهام معنایی از هم‌نگارها در سنتز متن به گفتار) به این مسئله می‌پردازد که در سنتز گفتار در زبان انگلیسی با واژه‌ها و اعدادی برخورد می‌کنیم که تلفظ آن‌ها بدون بافت قابل تشخیص نیست. وی چنین هم‌نگارهایی را به هفت دسته تقسیم می‌کند (Yarowsky 1997):

۱. هم‌نگارهایی که مقوله اجزای واژگانی کلام آن‌ها با هم فرق دارد؛ مانند هم‌نگار «lives» در دو جمله زیر (که در جمله اول اسم و در جمله دوم فعل است):

*Three lives were lost (lives (Noun/ N))

*One lives to eat (lives (Verb/ V))

۲. هم‌نگارهایی که مقوله اجزای واژگانی کلام آن‌ها یکسان است، ولی تلفظ آن‌ها یکسان نیست؛ مانند «beis/»bass/ به معنای «پایین‌ترین سطح نت موسیقی» و «baes/»bass/ به معنای «یک نوع ماهی» که هر دو اسم هستند. یا «bau/»bow/ به معنای «تعظیم» و «bou/»bow/ به معنای «گره پایونی» که هر دو اسم هستند.

۳. اسامی خاص: واژه‌هایی مانند «Nice» و «Begin» هم می‌توانند اسم خاص باشند و هم صفت و فعل که در جایگاه ابتدای جمله یا عنوان و غیره (که حرف اول با حروف بزرگ است) این هم‌نگارها ایجاد می‌شوند.

۴. اعداد رومی: اعداد رومی در بافت‌های گوناگون تلفظ‌های مختلفی دارند؛ مانند Chapter III: /tʃæptə ˈθri/ و Henry III: /ˈhɛnri ˈθɜːrd/.

۵. اعشار و تاریخ‌ها: به‌عنوان مثال، در انگلیسی 5/16 می‌تواند «five-sixteenths» (پنج‌شانزدهم) یا «May 16th» (شانزدهم ماه می) تلفظ شود.

۶. سال‌ها و کمیت‌نماها^۱: در انگلیسی اعدادی مانند 1750 می‌تواند به صورت «seventeen-fifty» (سال هزار و هفتصد و پنجاه) تلفظ شود (زمانی که به عنوان تاریخ به کار می‌روند) و یا به صورت «one-thousand seven-hundred fifty» (هزار و هفتصد و پنجاه) تلفظ شود (زمانی که قبل از واژه‌های مربوط به واحدهای اندازه‌گیری به کار روند؛ مانند «1750 miles»؛ (البته در فارسی هر دو حالت تلفظ یکسان دارد).

۷. مخفف‌ها^۲: مخفف‌ها ممکن است چندین تلفظ داشته باشند، مانند «St.» که می‌تواند مخفف «Saint» یا «Street» باشد و «Dr.» که می‌تواند مخفف «Doctor» یا «Drive» باشد.

در زبان فارسی نیز «بی‌جن‌خان و مرادزاده» هم‌نگاره‌های فارسی را به چند دسته تقسیم می‌کنند که از میان آن‌ها می‌توان به موارد زیر اشاره کرد (محسنی ۱۳۸۷ نقل از بی‌جن‌خان و مرادزاده ۱۳۸۳):

◇ واژگانی و نحوی: منظور از هم‌نگاره‌های واژگانی، هم‌نگاره‌هایی هستند که در فرهنگ‌های لغات به صورت مدخل‌های مجزا ذکر شده‌اند؛ مانند «شیر» و «شیر». حضور هم‌نگاره‌های نحوی وابسته به بافت نحوی و اضافه شدن تکواژهای اشتقاقی و تصریفی است؛

◇ هم‌نگاره‌هایی که در اثر عدم نمایش علائم زیروزبری در خط فارسی ایجاد شده‌اند؛ مانند: «مرد» /mard/ و «مرد» /mord/، «کند» /kanad/، «کند» /kond/ و «کند» /konad/، «سبک» /sabk/ و «سبک» /sabok/، «عرفا» /orfan/ و «عرفا» /orafa/، «عقلا» /oGalb/ و «عقلا» /aGlan/، «کمی» /kammi/ و «کمی» /kami/

◇ هم‌نگاره‌هایی که در اثر عدم تناظر یک-به-یک میان واج‌ها و نگاره‌ها در فارسی ایجاد شده‌اند؛ مانند «رود» /rud/ و «رود» /ravad/، «شوم» /f um/ و «شوم» /f avam/

◇ هم‌نگاره‌هایی که در اثر یکسانی تظاهر واجی و نوشتاری تکواژها به وجود می‌آیند؛ مانند یکسان بودن نمود نوشتاری تکواژ یای نکره، یای اسم‌ساز (اسم مکان، اسمی که دال بر شغل یا محافظت و دارندگی است، اسم معنی یا اشیا، تصغیر و تحبیب، اسم مصدر یا حاصل مصدر)، شناسهٔ دوم شخص مفرد و یای صفت‌ساز (صفت فاعلی و مفعولی، صفتی که دال بر نسبت است) که همه نمود نوشتاری «-ی» را دارند. یکسانی تکواژهای ضمیر متصل سوم شخص مفرد با تکواژ اسم‌ساز («-ش»). مانند:

ضمیر متصل سوم شخص مفرد: به رویش خندیدم. /rujaʃ/ و تگواژ اسم‌ساز: رویش گل‌ها را بین. /ruʃej/

- ◇ هم‌نگاره‌هایی که مربوط به رابطه میان وزن کلمات عربی و بعضی پسوند‌های فارسی می‌شوند. به‌عنوان مثال، کلمه «منزلت» /manzalat/ در باب مفعله، معنایی متفاوت از کلمه «منزلت» /manzalat/ با تعبیر «منزل تو»، دارای پی‌چسب ملکی «ت» /-at/ دارد؛
- ◇ هم‌نگاره‌های تکیه‌ای و غیرتکیه‌ای: تعداد این هم‌نگاره‌ها بسیار کم است. از میان آن‌ها می‌توان به هم‌نگاره‌های «ولی» /va'li/ به معنای «سرپرست»، «ولی» /'vali/ به معنای «اما» و «گویا» /gu'jo/ به معنای «واضح»، «گویا» /'gujo/ به معنای «مثل اینکه» اشاره کرد.

در راستای رفع ابهام از واژگان مبهم، پژوهشگران مدل‌های گوناگونی برای زبان‌های گوناگون، از جمله فارسی ارائه کرده‌اند که از میان آن‌ها می‌توان به موارد زیر اشاره کرد:

«مسعودی و راحتی فوجانی» مدلی برای رفع ابهام از واژگان مبهم فارسی با استخراج ویژگی‌های جدید پیشنهاد داده‌اند. برای ایجاد این مدل، دو دسته ویژگی واژگان و نشانه‌های همراه واژه مبهم و ویژگی‌هایی که با به‌کار بردن روش‌های مدل‌سازی موضوع به‌دست می‌آید، استفاده شده است. مدل موضوعی، مدلی آماری برای استخراج چکیده موضوع‌های موجود در اسناد یک پیکره است. آن‌ها در پژوهش خود از روش بدون سرپرست تخصیص پنهان «دریکله»^۱ (LDA) برای این منظور استفاده کرده‌اند. نتایج آزمایش‌ها برای پانزده واژه مبهم پرتکرار در زبان فارسی که از پیکره پژوهشکده پردازش هوشمند علائم استخراج شده، دقت حدود ۹۷ درصد را نشان می‌دهد که بیانگر تأثیر این روش در یافتن معنای مناسب واژگان مبهم است (۱۳۹۴).

«علایی» به معرفی رویکردی ماشینی برای چگونگی رفع ابهام معنایی از کلمات می‌پردازد. در زبان فارسی، که ساخت‌واژه پیچیده دارد، هم‌نگاره‌های بسیاری ساخته می‌شوند که معانی گوناگونی در بافت‌های گوناگون دارند. یکی از راه‌هایی که کمک می‌کند رفع ابهام از معنای کلمات مبهم (هم‌نگاره‌ها) با

1. latent Dirichlet allocation (LDA)

سهولت و دقت بیشتری انجام شود، تخصیص برچسب درست نحوی به کلمات است. بنابراین، اگر برچسب‌دهی نحوی قبل از مرحله رفع ابهام معنایی از کلمات، صورت پذیرد، رفع ابهام معنایی از هم‌نگاره‌ها با دقت بیشتری انجام خواهد گرفت. از آنجا که فراوانی هم‌نگاره‌های اسمی و صفتی در متون فارسی، در مقایسه با سایر هم‌نگاره‌ها بالاست، پس از تخصیص برچسب نحوی به کلمات، لازم است رفع ابهام از برچسب نحوی هم‌نگاره‌ها نیز صورت گیرد. در این مقاله، ابتدا روش‌های ماشینی موجود در جهت رفع ابهام از معنای کلمات معرفی شده است و سپس، الگوریتم «لسک»^۱ (که یکی از روش‌های یادگیری ماشینی بدون نظارت/ بدون سرپرست برای رفع ابهام معنایی از کلمات مبهم موجود در متون گوناگون است)، معرفی شده و سرانجام، رویکردی ماشینی برای رفع ابهام از معنای کلمات با استفاده از نتیجه مرحله برچسب‌زنی نحوی به کلمات و رفع ابهام از برچسب نحوی کلمات و الگوریتم «لسک» معرفی شده است. انجام برچسب‌دهی نحوی و رفع ابهام از برچسب نحوی هم‌نگاره‌ها باعث می‌شود که الگوریتم «لسک» تنها معانی مرتبط با برچسب‌های نحوی را در رفع ابهام معنایی از کلمات در نظر گیرد و در نتیجه، عمل رفع ابهام از معنای کلمات با دقت و سهولت بیشتری انجام پذیرد (۱۳۹۷).

«خیرمند پاریزی و نورمندی‌پور» نیز از میان چهار رویکرد متفاوت برای رفع ابهام معنایی کلمات (رویکرد مبتنی بر دانش، رویکرد تحت نظارت/ باسرپرست، رویکرد نیمه-نظارتی، و رویکرد بدون نظارت/ بدون سرپرست)، روش یادگیری مبتنی بر نمونه^۲ را که بر اساس رویکرد نظارتی است، برگزیده‌اند. در پژوهش آن‌ها دو دسته ویژگی مبتنی بر کلمات موضوعی معرفی شده است: دسته اول، وجود و نبود کلمات موضوعی را وابسته به وزن هر کلمه موضوعی تعیین می‌کند و دسته دوم، مجموع وزن کلمات موضوعی در هر جمله را مشخص می‌سازد. بعد از انجام عملیات پیش‌پردازش روی نمونه‌های ذخیره‌شده برای هر واژه، دو ماتریس ویژگی استخراج می‌شود. هر ماتریس ویژگی به عنوان ورودی نرم‌افزار «وکا»^۳ در نظر گرفته شده است. سپس، با استفاده از طبقه‌بندی کننده مبتنی بر

1. Lesk

2. Instance Based Learning (IBL)

3. Weka

نمونه و با روش اعتبارسنجی متقابل ۱۰ تایی نتایج را ارائه داده‌اند. سرانجام، نتایج حاصل از هر ماتریس ویژگی برای یافتن ویژگی بهینه مورد بررسی قرار گرفته است. همچنین، روش پیشنهادی آن‌ها روی سایر الگوریتم‌های یادگیری ماشین مورد بررسی قرار گرفته و نتایج ارائه شده است. میانگین صحت عملکرد روش پیشنهادی ۸۸/۳۱ درصد گزارش شده است (۱۳۹۵).

«علی‌پوری حافظی، مولودی و بیات» از رویکرد غیرنظارتی/بدون سرپرست برای رفع ابهام استفاده کرده‌اند. هدف پژوهش آن‌ها، بررسی امکان رفع ابهام معنایی از واژگان هم‌آوا-هم‌نویسه (هم‌نگاره) زبان فارسی با بهره‌گیری از هماندهای آنان بوده است که این امکان بر اساس یافته‌های پژوهش وجود دارد و حتی در مورد برخی از واژگان، احتمال بسیار بالایی را برای این منظور می‌توان ارائه کرد (۱۳۹۸).

«دلجوان انوری» بر این باور است که اگرچه سیستم‌های مبتنی بر پیکره و نظارت‌شده بهترین نتایج رفع ابهام را دارند، اما منابع دانش مورد نیاز این سیستم‌ها مانند پیکره‌های برچسب‌خورده معنایی، نیازمند هزینه‌های بسیار زیادی برای ساخت هستند. در نتیجه، انواع دیگری از سیستم‌های رفع ابهام، به‌ویژه برای زبان‌هایی که دسترسی به چنین منابعی وجود ندارد، نیاز داریم. در میان سیستم‌های رفع ابهام موجود، سیستم‌های مبتنی بر دانش و به‌ویژه روش‌های مبتنی بر گراف، از آنجا که به منابع محدود و دسترسی همچون فرهنگ لغات و هستان‌شناسی‌ها نیاز دارند، مورد توجه بسیاری از پژوهشگران در زمینه پردازش زبان طبیعی قرار گرفته‌اند. یکی از پیاده‌سازی‌های خوب مربوط به این روش‌ها در قالب ابزاری به نام UKB^۱ است و با استفاده از منابعی مانند پایگاه دانش «وُردنت» برای رفع ابهام مبتنی بر گراف کلمات زبان‌های انگلیسی، اسپانیایی و کاتالان انجام شده است. در این پژوهش، با در نظر گرفتن چالش‌های موجود در پردازش زبان فارسی، پیاده‌سازی‌هایی برای استفاده از این ابزار برای رفع ابهام کلمات زبان فارسی، مانند ساخت فرهنگ لغت و گراف از پایگاه دانش «فارس‌نت» انجام شده است. نتایج حاصل بر روی تعدادی از مجموعه داده‌های

1. Unsupervised Knowledge-Based (UKB)

آزمایشی فارسی، نشان‌دهنده کارایی پیاده‌سازی‌های انجام‌شده در این پژوهش است (۱۳۹۷).

«مظفری» و همکاران سعی کرده‌اند با ارائه الگوریتمی جدید مبتنی بر قالب‌های معنایی، سامانه‌ای قاعده‌مند برای رفع ابهام معنایی از حروف اضافه «از»، «در»، «با» و «تا» در زبان فارسی ارائه دهند. رفع ابهام معنایی از حروف اضافه، به‌خصوص در زبان فارسی، در پژوهش‌های مربوط به رفع ابهام معنایی همواره نادیده گرفته شده است. در پژوهش آن‌ها، ۱۰۰۰ جمله داده آموزشی، ۱۰۰ جمله داده توسعه و ۵۰۰ جمله داده تست مورد استفاده قرار گرفته است که از منابع اینترنتی و شبکه‌های اجتماعی مانند «یوتیوب»، جمع‌آوری گردیده است. برای انجام کار، تمام حروف اضافه مورد نظر پژوهش برچسب‌دهی معنایی شده و برای آن‌ها در پیکره آموزشی، قالب‌های معنایی بر اساس زبان قالب بنیاد (Minsky 1975) تعریف شده است. همچنین، برای کلمات قبل و بعد از حروف اضافه در پیکره نیز قالب‌هایی تهیه و در فایل داده مدخل‌ها وارد سامانه گردید. الگوریتم پژوهش در سه مسیر با استفاده از اطلاعات موجود در قالب‌ها، اقدام به تعیین معنای حروف اضافه در جمله‌ها می‌کند. نتایج آزمایش‌های داده تست، نشان‌دهنده دقت بالای عملکرد سامانه (۹۹/۱۶ درصد) در رفع ابهام معنایی از حروف اضافه در زبان فارسی است (۱۳۹۷).

«تویلی» و همکاران دو آزمایش مربوط به هم‌نگاره‌ها انجام داده‌اند. در آزمایش نخست، ۵۶۶ هم‌نگاره از ۱۹۲ آزمودنی گردآوری شده است. سپس، فراوانی رخداد سه معنای بسیار رایج همراه با وابسته‌های مربوطه و میزان ابهام کلی هر هم‌نگاره به این صورت گزارش شده است: هم‌نگاره‌هایی که معنای آن‌ها با توجه به تغییر در اجزای واژگانی کلام تغییر می‌کند، به‌طور کلی، مبهم‌تر از هم‌نگاره‌هایی هستند که تفاوت معنایی آن‌ها مربوط به یک طبقه دستوری می‌شود. از نظر میزان ابهام، هم‌نگاره‌هایی که تلفظ آن‌ها وابسته به معناست (هم‌آوا نیستند)، فرقی با آن‌هایی ندارند که هم‌آوا هستند. فراوانی واژه‌ها نیز ارتباطی به ابهام ندارد. سپس، توازن هم‌نگاره‌ها در پژوهش‌های گوناگون مقایسه شده و هم‌نگاره‌هایی که فراوانی نسبی هر دو معنای آن‌ها تقریباً مساوی است (هم‌نگاره‌های متوازن)، و هم‌نگاره‌هایی که یک معنای

غالب مشخص دارند (هم‌نگاره‌های قطبی شده) شناسایی شده‌اند. در آزمایش دوم، اعتبار دسته‌بندی معنایی برای زیرمجموعه‌ای از هم‌نگاره‌ها در آزمایش اول اندازه‌گیری شده و مشخص شده است که دسته‌بندی‌های معنایی از اعتبار بالایی برخوردار هستند (Twilley et al. 1994).

«یاروسکی» روش تصمیم‌گیری مبتنی بر آمار برای رفع ابهام معنایی در سنتز متن به گفتار معرفی می‌کند. به این صورت که بر اساس فهرست تصمیم‌گیری درباره معنای واژه مبهم الگوریتم، الگوهای نحوی نزدیک به هم / محلی^۱ را با شواهد باهم‌آیی دورتر واژه‌ها ترکیب می‌کند و قدرت درخت‌های تصمیم‌گیری را با برجسب گذارهای چندنگاشتی / چندتایی^۲ و طبقه‌بندی‌کننده‌های «بیزین»^۳ ترکیب می‌کند. الگوریتم روی هفت نوع اصلی از ابهام به کار گرفته شده است که در آن بافت برای انتخاب تلفظ واژه‌ها استفاده می‌شود (Yarowsky 1997).

در زمینه کاربرد الگوریتم‌ها و تکنیک‌های یادگیری ماشینی برای رفع ابهام معنایی از کلمات می‌توان به پژوهش انجام‌شده توسط «بکس» نیز اشاره کرد. وی به کاربرد بین پیکره‌ای سیستم‌های یادگیری ماشینی باسپرست برای رفع ابهام معنایی کلمات به منظور آزمودن توانایی تعمیم دادن در پیکره‌های متنی می‌پردازد. ابهام واژگانی در تمام زبان‌های طبیعی چالش‌برانگیز است و در زبان‌هایی که در نظام نوشتاری آن‌ها علائم زیروزبری (مانند علامت فتحه، کسره، تنوین، سکون) وجود دارد (مانند عربی)، اما این علائم معمولاً در خط حذف می‌شوند، بسیار رایج است. حذف علائم زیروزبری منجر به افزایش تعداد هم‌نگاره‌ها می‌گردد. نگه داشتن / نمایش علائم زیروزبری، به لحاظ نظری امکان‌پذیر است و به رفع ابهام هم‌نگاره‌ها کمک می‌کند، اما در عمل منجر به کاهش عملکرد پردازش زبان می‌شود (Bakx 2006).

«القحطانی، الدرماکی و دیاب» نیز رویکردهایی را برای ایجاد خودکار زیرمجموعه‌ای از کلمات به منظور حفظ علائم زیروزبری پیشنهاد می‌کنند که منجر به رفع ابهام معنایی هم‌نگاره‌ها می‌شود. آن‌ها در مقایسه با حالتی که همه علائم زیروزبری حذف یا حفظ شده باشند، رویکردهای پیشنهادی پایگاه

1. local

2. N-gram taggers

3. Bayesian classifiers

دادگانی را فراهم می‌کنند که به‌طور انتخابی علائم زیروبری همراه واژه‌ها به‌کار رفته است. سرانجام، می‌توان گفت که رویکرد پیشنهادی به‌عملکرد متوازن‌تر و منسجم‌تر منجر می‌شود (Alqahtani, Aldarmaki & Diab. 2019).

۳. روش پژوهش

استخراج و دسته‌بندی هم‌نگاره‌ها

برای استخراج هم‌نگاره‌ها از پیکره متنی فارسی، واژه‌هایی که بیش از یک برچسب اجزای واژگانی کلام داشتند، از پیکره استخراج شدند. شایان ذکر است که برچسب‌ها نیز به‌صورت مختصر هستند؛ به این مفهوم که تنها مقوله‌های اصلی (مانند اسم، فعل، صفت، حروف اضافه و غیره) را نشان می‌دهند؛ مانند نمونه زیر:

	AJ:3	N:29	ژیونی
	AJ:1	N:4	زحل
	N:3	AJ:66	زحمتکش
	AJ:1	N:189	زخم
	AJ:1	N:1	زخم خورده ی
	AJ:1	N:2	زخمه
ADV:1	N:84	AJ:384	زخمی
	V:2	N:742	زدن
	AJ:2	V:717	زده
	AJ:4	N:88	زر
	N:3	AJ:6	زراعی
	N:17	AJ:28	زرتشی
	AJ:1	N:1	زرخرید
	N:19	AJ:195	زرد
	ADV:1	AJ:1	زردتر
	N:5	AJ:28	زردشی
	AJ:2	N:17	زردی
	CL:1	N:9	زرع
	AJ:1	N:4	زركوب
	N:1	AJ:18	زرنگ
	AJ:1	N:7	زرنگار
	AJ:2	N:8	زرنگ.

و هم با جزئیات بیشتر برچسب‌ها (مانند اینکه یک فعل چه شناسه و چه زمانی دارد، یا اسم مفرد است یا جمع)؛ مانند نمونه زیر:

N,COM,SING,TIME:1	N,COM,SING,EZ:3	AJ,SIM:10	N,COM,SING:72	ابد
	RES,FW,ARAB:5	ADV,GENR,NEGG:11	ADV,GENR,SIM:37	ابدا
		N,COM,SING,EZ:72	N,COM,SING:147	ابداع
		N,COM,PL:9	N,COM,PL,EZ:18	ابداعات
	N,COM,SING,YA:3	AJ,SIM,EZ:5	AJ,SIM:7	ابدایی
		N,PR,SING:1	N,PR,SING,LOC:1	ابدانه
ADV,GENR,SIM:1	N,PR,SING:1	AJ,SIM,EZ:16	AJ,SIM:78	ابدی
		N,COM,SING,EZ:4	N,COM,SING:29	ابدیت
		N,COM,SING,EZ:58	N,COM,SING:148	ابر
		N,COM,SING:1	N,COM,SING,EZ:1	ابراء
N,PR,PL:4	N,PR,SING:12	N,COM,SING:17	N,COM,PL:30	ابرار
		N,COM,SING:299	N,COM,SING,EZ:1227	ابرارز
		AJ,SIM,EZ:1	AJ,SIM:4	ابرارزنده
		N,COM,SING,EZ:2	N,COM,SING:5	ابرام
		N,PR,SING:1	N,COM,SING:4	ابرانسان
	RES,FW,ARAB:7	N,PR,SING,EZ:211	N,PR,SING:295	ابراهیم
		N,PR,SING,EZ:3	N,PR,SING:10	ابراهیم‌زاده
	N,PR,SING,EZ:8	AJ,SIM:16	N,PR,SING:137	ابراهیمی
		N,PR,SING,LOC,EZ:1	N,PR,SING:2	ابردین
		N,COM,SING,EZ:3	N,COM,SING:4	ابردایانه

در پژوهش حاضر، به‌منظور بررسی و دسته‌بندی هم‌نگاره‌ها، استخراج فهرستی که در آن برچسب‌های اصلی (نه با تمام جزئیات) کنار واژه‌ها درج شده باشد، کافی بود. بنابراین، از فهرستی که حاوی ۱۴ برچسب اصلی معرفی شده در (Bijankhan et al. 2011) بود، استفاده شد (جدول ۱).

جدول ۱. فهرست برچسب‌های پیکره متنی فارسی

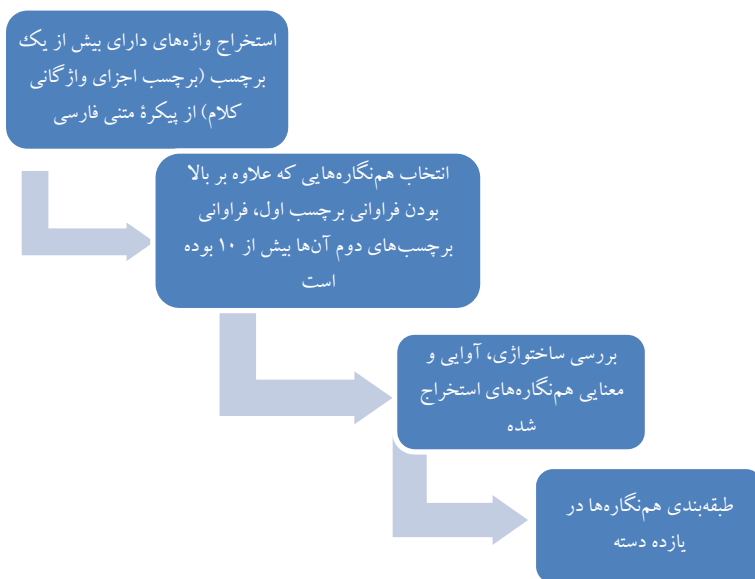
مثال	Tag Name	POS tag
کشاورز، خانه، کتاب	Noun	N
از، در، برای	Preposition	PREP
نقطه، ویرگول، علامت سؤال	Punctuation	PUNC
زیبا، اجتماعی، بزرگ	Adjective	AJ
می‌تواند، خورد، شمرد	Verb	V
که	Conjunction	CON
۵۰، ۹۰، ۱۲	Number	NUM
من، تو، ایشان	Pronoun	PRO
این، آن، هر	Determiner	DET

مثال	Tag Name	POS tag
یقیناً، خوب، بسیار	Adverb	ADV
را	Postposition	POSTP
هر برچسبی غیر از برچسب‌های اصلی	Residual	RES
نوع، دست، چنین	Classifier	CL
ای، یا	Interjection	INT

فهرست استخراج شده شامل ۱۰۹۷۸ واژه است که بیش از یک برچسب اجزای واژگانی کلام دارند. برچسب‌ها در هر ردیف کنار هر واژه مشخص شده‌اند. در مرحله بعد به بررسی فراوانی برچسب‌های هر ردیف پرداخته شد. به این صورت که ستون اول، اولین برچسب هر واژه را نشان می‌دهد که قاعده‌تاً فراوانی بیشتری نسبت به سایر برچسب‌های تخصص داده‌شده به آن واژه دارد. به ترتیب، هرچه از ستون اول از سمت راست به سمت چپ حرکت می‌کنیم، فراوانی برچسب‌ها کمتر می‌شود. به‌طور کلی، می‌توان گفت که برچسب‌های اول و دوم هر واژه در مقایسه با سایر برچسب‌های تخصیص داده‌شده به هر واژه، فراوانی بیشتری داشتند. بنابراین، برای انتخاب هم‌نگاره‌ها، فراوانی برچسب‌های اول و دوم مورد بررسی دقیق قرار گرفت. فراوانی برچسب‌های اول هم‌نگاره‌ها به‌طور متوسط، بین ۱۰ تا چند هزار می‌رسد؛ در حالی که فراوانی برچسب‌های دوم به‌ندرت به بالاتر از ۱۰۰ می‌رسد. نگارنده، فراوانی حداقل ۱۰ را برای برچسب دوم در نظر گرفت و بر همان اساس، هم‌نگاره‌ها را از فهرست اول استخراج کرد. به نظر نگارنده، به‌عنوان نمونه، اگر فراوانی برچسب اول یک هم‌نگاره بسیار بالا باشد، اما فراوانی برچسب دوم آن زیر ۱۰ باشد (مثلاً ۲)، ارزش بررسی زبانی را ندارد، زیرا تنها دو نمونه در پیکره برای آن برچسب وجود دارد، در حالی که تعداد نمونه‌ها در بررسی‌های زبانی مسئله‌ای مهم است. شایان ذکر است که نگارنده به این امر واقف است که ممکن است برای پژوهشگر دیگری حتی یک نمونه هم حائز اهمیت باشد و در این حالت نیازی به استخراج فهرست دیگری از فهرست اول نیست. در هر صورت در این پژوهش، بر اساس فراوانی بیشتر از ۱۰ برچسب دوم، فهرست دومی از فهرست اولیه استخراج شد که شامل ۱۶۷۵ هم‌نگاره (واژه‌هایی با بیش از یک برچسب POS) است (پیوست ۱).

در گام بعدی، به بررسی ماهیت ساخت‌وازی و آوایی هم‌نگاره‌های استخراج شده از

پیکره متنی فارسی (فهرست استخراج شده از فهرست همه هم‌نگاره‌های به‌دست آمده از پیکره متنی فارسی) پرداخته شد تا بتوان دسته‌بندی صوری و آوایی مبتنی بر پیکره از هم‌نگاره‌های فارسی ارائه داد. مراحل طی شده برای دسته‌بندی پیکره-بنیاد هم‌نگاره‌های فارسی در نمودار جریان ۱، قابل مشاهده است:



شکل ۱. نمودار جریانی (block diagram) مراحل طبقه‌بندی پیکره‌بنیاد هم‌نگاره‌های فارسی

نتایج بررسی شامل دسته‌بندی‌های زیر است که به ترتیب، میزان فراوانی هر دسته ارائه شده است (شایان ذکر است که مثال‌ها و فراوانی آن‌ها از پیکره متنی فارسی استخراج شده، اما کاربرد آن‌ها در عبارت‌ها و جمله‌هایی که برای نمونه ارائه شده‌اند، برگرفته از ششم زبانی نگارنده است):

- دسته اول (۶۳۱ مورد در پیکره، ردیف ۱ تا ۶۳۱ در پیوست): هم‌نگاره‌هایی هستند که با اضافه شدن تکواژ «ی» به ستاک ایجاد شده‌اند؛ مانند «ابزاری»، «اتفاقی»، «احساساتی»، «احساسی»، «اختیاری»، «پزشکی»، «پرداختی»، «توافقی»، «جهانی»، «خارجی»، «رقمی»، «زمانی»، «ساختاری»، «شهری»، «غذایی»، «فردی»، «کتابی»، «مذهبی»، «مردمی»، «نژادی»، «هنری»، «هوایی» و مواردی از این دست. این هم‌نگاره‌ها خود به‌صورت زیر دسته‌بندی می‌شوند:

◇ واژه‌بست «-ی» یعنی «یای نکره» در نقش حرف تعریف نامعین در پایان اسم (و در نوشتار، متصل به آن)، یا اسم عدد «یک» پیش از اسم، و یا هر دو با هم به کار می‌رود (لازار ۱۳۸۹ ترجمه مهستی بحرینی). واژه‌بست، تکواژ دستوری وابسته فاقد تکیه است که در ساخت‌های بزرگ‌تر از واژه شرکت می‌کند و از نظر آوایی با پایه خود یک واژه واجی می‌سازد. واژه واجی در واقع، یک گروه یا ساخت نحوی است که از نظر آوایی همانند یک واژه تلفظ می‌شود، اما از نظر صرفی، نحوی و معنایی، یک واژه به شمار نمی‌آید (شقاقی ۱۳۹۱). همان‌گونه که اشاره شد، «-ی» نکره یکی از واژه‌بست‌ها محسوب می‌شود. از میان مثال‌های انبوهی که در پیکره وجود دارند و در این دسته قرار می‌گیرند، می‌توان واژه «اصلی» را نام برد. این واژه در ۳۹۲۳ مورد به صورت صفت /A/ به کار رفته است و در ۵۹ مورد به صورت اسم /N/ به کار رفته است. کاربرد این واژه در جایگاه صفت که بیشترین فراوانی را برای این واژه داشته است، در مواردی همچون «حساب اصلی»، «درب اصلی» و غیره می‌توان در نظر گرفت. اما یکی از کاربردهای این واژه در جایگاه اسم می‌تواند به صورت اسم مختوم به یای نکره باشد، مانند «اصلی کاربردی». پسوند «-ی» اشتقاقی دال بر مکان (صادقی ۱۳۷۱ الف): به عنوان مثال، در پیکره، واژه «نزدیکی» در ۶۷۵ مورد به صورت اسم (بیشتر اسم مکان) به کار رفته است که می‌توان در ساخت‌هایی مانند «در نزدیکی ما زندگی می‌کنند» چنین جایگاهی را برای این واژه در نظر گرفت. در ۷۰ مورد نیز در جایگاه صفت به کار رفته است؛ مانند «راه نزدیکی/ نزدیک‌تری می‌شناسی؟». در یک مورد هم در جایگاه فعل /V/ که حتماً در گونه گفتاری زبان به کار رفته است، مانند «نزدیکی؟» به معنای «آیا نزدیک هستی؟» (در چنین مواردی پسوند «-ی» علامت شناسه دوم شخص مفرد است).

پسوند اشتقاقی «-ی» دال بر محافظت، دارندگی و شغل (صادقی ۱۳۷۱ ب): نمونه‌های فراوانی در پیکره در این دسته قرار می‌گیرند. به عنوان مثال، در پیکره، واژه «دانشگاهی» در ۶۴۳ مورد به صورت صفت به کار رفته است که می‌توان در ساخت‌هایی مانند «مدرک دانشگاهی» چنین جایگاهی را برای این واژه در نظر گرفت. در ۵۳ مورد نیز در جایگاه اسم به کار رفته است که یکی از موارد کاربرد این واژه در جایگاه اسم، اسمی است که دال بر شغل است، مانند «دانشگاهی هستیم». البته، کاربرد این واژه در جایگاه اسم می‌تواند، به صورت اسم مختوم به

یای نکره هم باشد، مانند «دانشگاه تهران، دانشگاهی است که در آن ...».

پسوندهای اشتقاقی «-ی» دال بر اسم اشیا و اسم معنی (صادقی ۱۳۷۱ج): از میان مثال‌های انبوهی که در پیکره موجود هستند، می‌توان به واژه «سازمانی» اشاره کرد. این واژه در ۲۶۰ مورد به صورت صفت به کار رفته است که می‌توان در ساخت‌هایی مانند «رفتار سازمانی» چنین جایگاهی را برای این واژه در نظر گرفت. در ۱۲۲ مورد نیز در جایگاه اسم به کار رفته است که یکی از موارد کاربرد این واژه در جایگاه اسم، اسم معنی است؛ مانند «ساختار سازمانی».

پسوندهای اشتقاقی «-ی» دال بر مصدر و حاصل مصدر. این پسوند در دستورهای زبان فارسی یای حاصل مصدر یا یای مصدری نامیده شده است، اما ایران‌شناسان غربی آن را پسوند اسم معنی^۱ نامیده‌اند. بیشتر دستورنویسان ایرانی و هندی تنها نقش این پسوند را ساختن حاصل مصدر (= اسم حالت یا اسم معنی) از اسم و صفت دانسته‌اند (صادقی ۱۳۷۱ج). به عنوان مثال، در پیکره، واژه «خوبی» در ۱۱۹۹ مورد به صورت صفت به کار رفته است که می‌توان در ساخت‌هایی مانند «کتاب‌های خوبی» چنین جایگاهی را برای این واژه در نظر گرفت (که البته، در اینجا یای نکره به گروه اسمی متشکل از اسم + صفت «کتاب خوب» اضافه شده است). در ۸۵ مورد نیز در جایگاه اسم به کار رفته است که یکی از موارد کاربرد این واژه در جایگاه اسم، اسم معنی است؛ مانند «خوبی کردن به خلق». در سه مورد هم در جایگاه فعل به کار رفته است؛ نوعاً بیشتر در گونه گفتاری زبان، مانند «خوبی؟» به معنای «آیا خوب هستی؟» (در چنین مواردی پسوند «-ی» علامت شناسه دوم شخص مفرد است).

پسوندهای اشتقاقی «-ی» دال بر صفت فاعلی و مفعولی (صادقی ۱۳۷۲الف): به عنوان نمونه، در پیکره، واژه «تولیدی» در ۹۷۸ مورد به صورت صفت به کار رفته است که می‌توان در ساخت‌هایی مانند «کار تولیدی» چنین جایگاهی را برای این واژه در نظر گرفت. در ۴۰ مورد نیز در ساخت‌هایی مانند «تولیدی‌ها سهم بالایی در بهبود وضعیت صادرات دارند» در جایگاه اسم به کار رفته است.

پسوندهای اشتقاقی «-ی» / یای نسبت: این پسوند زایاترین و فعال‌ترین پسوند زبان فارسی است و به پایه‌های مختلف متصل می‌شود و دلالت بر معانی گوناگونی

1. abstract suffix

دارد؛ از جمله، مفهوم هیئت و شکل؛ مانند: استوانه‌ای، قلمی، کتابی، بادامی، آردی. مفهوم ماهیت؛ مانند: آجری، سنگی، آهنی، پوستی. مفهوم اتصاف و دارندگی؛ مانند: سمی، گوشتی، شکری، خجالتی، چادری، عینکی، موتوری. مفهوم وابستگی؛ مانند: ایرانی، تهرانی، شهری، روستایی، بغلی، صنعتی، زمینی، ورزشی، خوردنی، زبانی، اطلاعاتی، اسلامی، قانونی، مقرراتی. مفهوم شایستگی، الزام و احتمال؛ مانند: خوردنی، نوشیدنی، تعریفی، تماشائی، افزودنی، آوردنی، عملی. مفهوم فاعلیت و مفعولیت (که پیش از این توضیح داده شد)؛ مانند: تولیدی، اجرائی، فراری، ورودی، پختنی، حاضری، اعدامی، بافتنی. مفهوم طریقه که پایه، وسیله و طریقه‌ای را نشان می‌دهد که فعل جمله به آن طریقه بر آن واقع می‌شود یا نحوه‌ای که موصوف با آن وزن، عرضه، فروخته، استخدام، اجاره یا انجام می‌شود؛ مانند: زمینی، تلفنی، شفاهی، کارتی، بلیطی، مکاتبه‌ای، جعبه‌ای، قالبی، دانه‌ای، رسمی، قراردادی، پیمانی، نوبتی، تصادفی، (صادقی ۱۳۷۲ ب). مثال‌های فراوانی در پیکره در این دسته‌بندی قرار می‌گیرند؛ از میان آن‌ها می‌توان به واژه «شهری» اشاره کرد که در ۱۲۸۷ مورد به‌صورت صفت به کار رفته است و می‌توان چنین کاربردی را در ساخت‌هایی مانند «زندگی شهری» تصور کرد. در ۲۴۸ مورد هم در جایگاه اسم به کار رفته است که بیشتر یای نکره متصل به اسم است؛ مانند «تهران، شهری با جمعیت ...».

۲. دسته دوم، ۱۷۰ مورد در پیکره (ردیف ۶۳۳ تا ۸۰۳ در پیوست): هم‌نگاره‌هایی هستند که یا واژه‌های وام گرفته از عربی هستند یا بر اساس قالب‌های عروضی عربی (بیشتر مفعول، فاعل، مفعول، مفاعل، افتعال، متفعل و غیره) ساخته شده‌اند و در جایگاه اسم، صفت، قید و غیره به کار رفته‌اند. هم‌نگاره‌هایی مانند «مانع»، «متجاوز»، «متفکر»، «محصل»، «مخلوق»، «علیه»، «خالص»، «منکر»، «ذیل»، «فوق»، «قبل»، «صحیح»، «منقول»، «واقف»، «مهاجم»، «مجروح»، «متقاضی»، «ضد»، «شی / شیء»، «صبح»، «شهید»، «مکرر»، «مقابل» و «مستند» در این دسته قرار می‌گیرند؛

۳. دسته سوم، ۱۲۹ مورد در پیکره (ردیف ۸۰۵ تا ۹۳۴ در پیوست): هم‌نگاره‌هایی هستند که هم‌آوا نیستند و در اثر عدم نمایش علائم زیروبری در خط فارسی و عدم تناظر یک-به-یک میان واج‌ها و نگاره‌ها در فارسی و یا تغییر در تکیه به‌وجود آمده‌اند. همان‌گونه که در مقدمه بیان شد، «بی‌جن‌خان و مرادزاده» این هم‌نگاره‌ها را در سه

دسته قرار داده‌اند (محسنی ۱۳۸۷ نقل از بی‌جن خان و مرادزاده)، اما به نظر نگارنده همهٔ مواردی را که در این سه دسته قرار می‌گیرند، می‌توان در یک دستهٔ کلی‌تر تحت عنوان «هم‌نگاره‌های غیرهم‌آوا» قرار داد. نمونه‌های فراوانی در پیکرهٔ متنی فارسی برای این دسته می‌توان یافت که از میان آن‌ها می‌توان به این موارد اشاره کرد: «آرد» با دو تلفظ /ɔrd/ (۱۴۳ مورد اسم) و /brad/ (۱۱ مورد فعل)، «بیر» با دو تلفظ /babr/ (۳۸ مورد اسم) و /bebar/ (۱۹ مورد فعل)، «پر» با دو تلفظ /por/ (۸۶۸ مورد صفت و ۴ مورد قید) و /par/ (۸۱ مورد اسم)، «سبک» با دو تلفظ /sabk/ (۴۳۰ مورد اسم) و /sabok/ (۲۳۴ مورد صفت و ۲ مورد قید)، «زند» با دو تلفظ /zanad/ (۸۰ مورد فعل) و /zand/ (۸۰ مورد اسم)، «مردم» با دو تلفظ /mardam/ (با معنی «مرد هستم» (اسم+فعل) ۱۳ مورد) و /mardom/ (۱۶۱۲۴ مورد اسم، که البته موارد اسم می‌توانسته به معنای «مرد من» (اسم+ضمیر ملکی) هم بوده باشد که در این صورت تلفظ /mardam/ را دارد)، «کرد» با دو تلفظ /kard/ (۳۳۴۵۳ مورد فعل) و /kord/ (۷۳ مورد صفت و ۵۵ مورد اسم)، «کند» با دو تلفظ /kanad/ (۱۱۸۸۶ مورد فعل) و /kond/ (۱۴۰ مورد صفت، ۱۳ مورد قید و ۴ مورد اسم)، «چکیده» که تغییر تکیه دو تلفظ را ایجاد کرده است: /tʃekide/ (۵۰ مورد اسم و ۲۴ مورد صفت) و /tʃekide/ (۱۹ مورد فعل)، «ماشینی» که تغییر تکیه دو تلفظ را ایجاد کرده است: /moʃini/ (۷۹ مورد صفت) و /moʃini/ (۱۴ مورد اسم)، «گویا» که تغییر تکیه دو تلفظ را ایجاد کرده است: /gujɔ/ (۲۲۷ مورد قید) و /guʃɔ/ (۳۳ مورد صفت و ۶ مورد اسم)، «رود» با دو تلفظ /rud/ (۴۲۳ مورد اسم) و /ravad/ (۱۶۹ مورد فعل)، «نبرد» با دو تلفظ /nabard/ (۳۸۵ مورد اسم) و /nabarad/ (۴۰ مورد فعل)، «نشسته» با دو تلفظ /neʃaste/ و /naʃoste/ (۴۳۰ مورد فعل و ۵۰ مورد صفت و یک مورد اسم)، «تو» با دو تلفظ /to/ (۳۴۹۷ مورد ضمیر) و /tu/ (۱۴۵ مورد حرف اضافه، ۲۲ مورد اسم و ۵ مورد قید)، «درست» با دو تلفظ /dorost/ (۱۲۵۱ مورد صفت و ۶۷۷ مورد قید) و /darsat/ (۱۴ مورد اسم به معنی «درس تو»، «ده» با دو تلفظ /dah/ (۱۳۷۰ مورد عدد) و /deh/ (۹۷ مورد اسم و ۴۳ مورد فعل به معنی «بده»؛

۴. دستهٔ چهارم، ۵۴ مورد در پیکره (ردیف ۹۳۶ تا ۹۹۰ در پیوست): اسامی خاص عربی یا فارسی هستند که به صورت صفت، قید و غیره هم به کار رفته‌اند؛ مانند «اشرف»، «آدم»، «آزاده»، «حافظ»، «شیرین»، «سوزان»، «رئوف»، «سلیم»، «سیما»، «سعید»، «مسعود»، «امین»، «یگانه» و «کبیر»؛

۵. دسته پنجم، ۵۲ مورد در پیکره (ردیف ۹۹۲ تا ۱۰۴۴ در پیوست): هم‌نگاره‌هایی هستند که با اضافه شدن پسوند «ه» به بن ماضی ساخته شده‌اند. گاهی ممکن است بعد از ساخته شدن، پیشوند منفی «نا» نیز به آن‌ها اضافه شده باشد. مواردی مانند «ایستاده»، «افزوده»، «نشسته»، «نادیده»، «مانده»، «بازنشسته»، «بسته»، «زاده»، «فرموده»، «خوانده»، «بریده»، «نامبرده» و «گمشده» که نوعاً برچسب فعل، اسم یا صفت دارند، در این طبقه قرار دارند. به عنوان مثال، هم‌نگاره «ایستاده» در عبارت «او دم در ایستاده و ...» دارای برچسب فعل و در عبارت «ایستاده میان گرمابه» برچسب صفت دارد؛ «نرسیده» در عبارت «او به گلستان نرسیده ...» برچسب فعل و در عبارت «میوه نرسیده»، برچسب صفت دارد؛

۶. دسته ششم، ۵۲ مورد در پیکره (ردیف ۱۰۴۶ تا ۱۰۹۸ در پیوست): هم‌نگاره‌هایی هستند که واحدهای شمارش و اندازه‌گیری هستند. مانند «لیتر»، «مگاوات»، «راس / رأس»، «پوند»، «جبه»، «تار»، «فقره»، «واحد» و «حلقه». این نوع هم‌نگاره‌ها بیشتر برچسب طبقه‌بندی کننده (Classifier/ CL) یا اسم دارند. به عنوان نمونه، هم‌نگاره «لیتر» در جمله «هر لیتر چند کیلوگرم است؟» برچسب اسم دارد و در جمله «۵۰ لیتر بنزین زدم»، دارای برچسب طبقه‌بندی کننده (CL) است؛

۷. دسته هفتم، ۴۴ مورد در پیکره (ردیف ۱۱۰۰ تا ۱۱۴۴ در پیوست): هم‌نگاره‌هایی هستند که به پسوندهای اشتقاقی «-مند»، «-گر»، «-گزار»، «-کار»، «-نده» ختم می‌شوند و نوعاً بیشتر اسم و صفت هستند. مانند «کارگر»، «آفریننده»، «دانشمند»، «درستکار»، «برگزارکننده»، «بازدارنده»، «هنرمند»، «خدمتگزار»، «کارگر»، «صادرکننده» و «گیرنده»؛

۸. دسته هشتم، ۳۶ مورد در پیکره (ردیف ۱۱۴۶ تا ۱۱۸۲ در پیوست): اسم‌هایی هستند که به عنوان افعال گذشته (سوم شخص مفرد)، حال (دوم شخص مفرد) و صفت نیز به کار می‌روند یا برعکس. مانند: «بازداشت»، «بازگشت»، «باد»، «بازیافت»، «برخاست»، «برگشت»، «بست»، «پرداخت»، «پیوست»، «ساز»، «نوش» و غیره؛

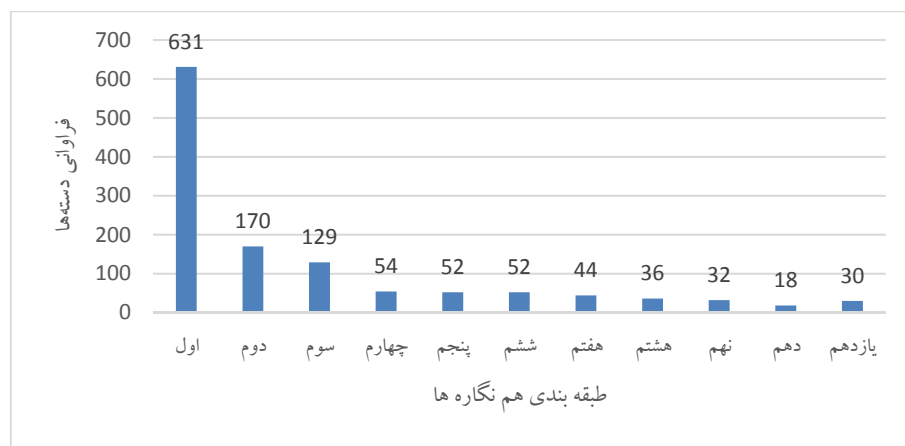
۹. دسته نهم، ۳۲ مورد در پیکره (ردیف ۱۱۸۴ تا ۱۲۱۶ در پیوست): مواردی هستند که به پسوند «-انه» ختم شده‌اند و نوعاً بیشتر قید و صفت و اسم هستند. هم‌نگاره‌هایی مانند «آگاهانه»، «سرانه»، «عجولانه»، «دیوانه»، «صادقانه»، «مردانه»، «میانه» و «هوشیارانه» در این دسته قرار می‌گیرند؛

۱۰. دسته دهم، ۱۸ مورد در پیکره (ردیف ۱۲۱۸ تا ۱۲۳۶ در پیوست): مواردی که عدد،

اعداد ترتیبی به‌تنهایی یا همراه با قید ماه و سال، و غیره هستند و در جایگاه اسم، قید، صفت و غیره به کار رفته‌اند. مانند «اول»، «یکبار»، «یکسره»، «سه‌ماهه»، «یکجا» و «سوم»؛

۱۱. دسته یازدهم، ۳۰ مورد در پیکره (ردیف ۱۲۳۸ تا ۱۲۶۸ در پیوست): این دسته شامل مواردی است که به ملیت، نژاد، اعتقاد مذهبی، ملی یا حزب خاصی دلالت دارند. این هم‌نگاره‌ها به‌طور معمول، در جایگاه اسم یا صفت به کار می‌روند. هم‌نگاره‌هایی مانند «ترکمن»، «صهیونیست»، «ایرانی»، «تاجیک»، «شیعه»، «سیاستمدار»، «سیاه‌پوست»، «فیلسوف»، «قاجار»، «مومن / مؤمن» و «تروریست» در این طبقه قرار می‌گیرند. به‌عنوان نمونه، هم‌نگاره‌های «ترکمن» و «تاجیک» در عبارت «ترکمن‌ها / تاجیک‌ها اقوامی هستند که...» دارای برچسب اسم و در عبارت «زن ترکمن / مرد ترکمن»، دارای برچسب صفت است. البته، این طبقه‌بندی معنایی است، به این مفهوم که در ظاهر واژه علامت ساخت‌واژی خاصی دیده نمی‌شود که بتوان قاعده‌ی صوری ارائه داد و بر اساس آن طبقه‌بندی انجام داد. شایان ذکر است که برخی از هم‌نگاره‌های مختوم به «ی» را نیز می‌توان در این دسته قرار داد (مانند «ایرانی»، «سنی» و...). اما چون این هم‌نگاره‌ها علامت صوری دارند، در همان دسته اول قرار گرفتند.

نمودار ۲، طبقه‌بندی پیکره-بنیاد هم‌نگاره‌های فارسی به همراه فراوانی هر دسته از هم‌نگاره‌ها را نشان می‌دهد.



نمودار ۲. طبقه‌بندی پیکره-بنیاد هم‌نگاره‌های فارسی و فراوانی هر دسته

شایان ذکر است که برخی از هم‌نگاره‌ها در بیش از یک طبقه قرار می‌گیرند. به‌عنوان نمونه واژه «ایرانی» هم در دسته اول (واژه‌های مختوم به «-ی») و هم در دسته یازدهم (ملیت) قرار دارد یا واژه «امین» هم در دسته اسم خاص قرار دارد و هم در دسته دوم (واژه‌هایی با قالب عروضی عربی).

هم‌نگاره‌های دیگری هم در پیکره متنی فارسی هستند که نمی‌توان در هیچ یک از ده دسته قرار داد و از نظر معنایی و ویژگی‌های ظاهری / ساخت‌واژی نیز نمی‌توان آن‌ها را دسته‌بندی کرد، اما به هر حال، در پیکره بیش از یک برجسب اجزای کلام دارند. این‌ها شامل این موارد می‌شود: «بست»، «بعضی»، «بلند»، «بن»، «پیر»، «تابناک»، «تازه»، «تیرانداز»، «خواست»، «خلوت» و «گذشت» و غیره (ردیف ۱۲۷۰ تا آخر ردیف آخر پیوست). بسیاری از این هم‌نگاره‌ها صفت‌هایی هستند که بسته به جایگاه‌شان در یک عبارت و جمله، می‌توانند به صورت صفت نیز به کار روند یا برعکس (مانند «بالا»، «باز»، «بهتر»، «همانند» «کم»، «سریع» «فوق‌العاده»، «بی تفاوت» و غیره). برخی دیگر صفت‌هایی هستند که به‌عنوان اسم نیز به کار می‌روند یا برعکس (مانند «تیرانداز»، «تئوری»، «تعطیل»، «حساب»، «خلاف»، «دستگیر»، «دوردست»، «راست» و غیره). نگارنده در این پژوهش، ۱۱ طبقه‌بندی از هم‌نگاره‌های فارسی بر اساس پیکره متنی فارسی ارائه داده است، اما به این امر آگاه است که ممکن است طبقه‌بندی‌های دیگری نیز بتوان ارائه داد یا از پیکره‌های دیگر فارسی نیز برای بررسی هم‌نگاره‌ها استفاده کرد.

۴. بحث و نتیجه‌گیری

در پژوهش حاضر به منظور بررسی و دسته‌بندی هم‌نگاره‌ها در یکی از پیکره‌های فارسی (پیکره متنی فارسی)، ابتدا واژه‌هایی که بیش از یک برجسب اجزای واژگانی کلام (POS) داشتند، از پیکره استخراج شد. فهرست استخراج‌شده شامل ۱۰۹۷۸ واژه بود. سپس، فراوانی برجسب‌های هر ردیف مورد بررسی قرار گرفت و فهرست دیگری استخراج شد که شامل هم‌نگاره‌هایی است که افزون بر فراوانی بالای برجسب اول واژه‌های آن، فراوانی برجسب دوم آن‌ها نیز حداقل ۱۰ بود. این فهرست شامل ۱۶۷۵ هم‌نگاره است. ماهیت ساخت‌واژی، آوایی یا حتی معنایی هم‌نگاره‌های استخراج‌شده از پیکره متنی فارسی بررسی شد و بر اساس آن هم‌نگاره‌های استخراج‌شده در یازده دسته طبقه‌بندی شدند.

از میان این هم‌نگاره‌ها، تنها هم‌نگاره‌های موجود در دسته یازدهم بر اساس معیار معنایی (ملیت، نژاد، اعتقاد مذهبی، ملی یا حزب خاص) طبقه‌بندی شده‌اند. سایر دسته‌ها بر اساس معیارهای ساخت‌واژی و آوایی طبقه‌بندی شده‌اند. شایان ذکر است که هم‌نگاره‌های دیگری نیز در فهرست استخراج‌شده قرار دارند که فراوانی قابل توجهی دارند، اما نمی‌توان معیار صوری یا معنایی برای طبقه‌بندی آن‌ها ارائه داد. این هم‌نگاره‌ها بیشتر، قیده‌ها، صفت‌های تعریف یا صفاتی هستند که به صورت اسم، صفت، قید، ضمیر و غیره نیز به کار می‌روند مانند: «اغلب»، «اینجوری»، «همین»، «همزمان»، «همان»، «درحقیقت»، «بهتر»، «خلاصه»، «بلندمدت»، «فوق‌العاده»، «پایین»، «تند»، «بسیار»، «بیشتر»، «بعضی» و موارد مشابه دیگر.

همان‌گونه که پیش‌تر نیز اشاره شد، «یاروسکی» هم‌نگاره‌های انگلیسی را به هفت دسته تقسیم کرده است که از میان آن‌ها سه دسته اول با پژوهش حاضر همپوشانی دارد: ۱- هم‌نگاره‌هایی که مقوله اجزای واژگانی کلام آن‌ها با هم فرق دارد، اما تلفظ یکسان دارند؛ ۲- هم‌نگاره‌هایی که مقوله اجزای واژگانی کلام آن‌ها یکسان است، اما تلفظ آن‌ها یکسان نیست؛ و ۳- اسامی خاص (Yarowsky 1997). در زبان فارسی نیز «بی‌جن‌خان و مرادزاده» هم‌نگاره‌های فارسی را به شش دسته تقسیم کرده‌اند (محسنی ۱۳۸۷ نقل از بی‌جن‌خان و مرادزاده) که از میان آن‌ها، دسته دوم (هم‌نگاره‌هایی که در اثر عدم نمایش علائم زیروزبری در خط فارسی، ایجاد شده‌اند؛ مانند «مرد» /mard/ و «مرد» /mord/)، دسته سوم (هم‌نگاره‌هایی که در اثر عدم تناظر یک-به-یک میان واج‌ها و نگاره‌ها در فارسی ایجاد شده‌اند؛ مانند «رود» /rud/ و «رود» /ravad/) و دسته ششم (هم‌نگاره‌های تکیه‌ای و غیرتکیه‌ای؛ مانند «ولی» /va'li/ و «ولی» /'vali/) با دسته سوم در پژوهش حاضر همپوشانی دارند، زیرا به نظر نگارنده سه دسته‌بندی «بی‌جن‌خان و مرادزاده» را می‌توان در یک دسته قرار داد: هم‌نگاره‌هایی که هم‌آوا نیستند و در اثر عدم نمایش علائم زیروزبری در خط فارسی و عدم تناظر یک-به-یک میان واج‌ها و نگاره‌ها در فارسی و یا تغییر در تکیه به وجود آمده‌اند. همچنین، بخشی از نمونه‌های موجود در دسته چهارم در طبقه‌بندی آن‌ها که مربوط به اضافه شدن پسوند «ی» به ستاک می‌شود، با دسته اول پژوهش حاضر همپوشانی دارد. شایان ذکر است که وجه تمایز پژوهش حاضر با پژوهش‌های انجام‌شده در زبان فارسی و انگلیسی این است که پژوهش حاضر پیکره-بنیاد است و تمام نمونه‌ها برگرفته از پیکره هستند که به ترتیب فراوانی مرتب شده‌اند. در

ضمن، استخراج هم‌نگاره‌ها از پیکره بر اساس تعداد برچسب‌های اجزای کلام اختصاص داده‌شده به هر واژه در پیکره بوده است.

حال که فهرست مبسوطی از هم‌نگاره‌های فارسی بر اساس پیکره به‌دست آمده (پیوست ۱)، می‌توان از رویکردهای رایج به‌منظور رفع ابهام از معنای کلمات مبهم/هم‌نگاره‌ها استفاده کرد. رویکردهای گوناگونی که برای این منظور به کار رفته و همه یا برخی از آن‌ها در زبان‌های گوناگون، به‌ویژه در زبان انگلیسی، به کار گرفته شده است، شامل این موارد است (Pal and Saha 2015):

◇ رفع ابهام مبتنی بر دانش^۱ (که از اطلاعات موجود در فرهنگ‌های لغات (وُردنت‌ها)، اصطلاحنامه‌ها و مانند آن استفاده می‌کند). این رویکرد خود چندین زیر شاخه دارد که شامل این موارد است: الگوریتم لسک^۲، تشابه معنایی^۳، ترجیحات انتخابی^۴ و روش کاوشی^۵؛

◇ رفع ابهام نظارتی/باسرپرست^۶ (که از روش‌های یادگیری ماشینی استفاده می‌کند). این رویکرد نیز خود می‌تواند شامل رویکردهای فهرست تصمیم^۷، درخت تصمیم^۸، ییز نایبو/ساده^۹، شبکه‌های عصبی^{۱۰}، یادگیری مبتنی بر نمونه^{۱۱} و ماشین برداری حمایتی^{۱۲} شود؛

◇ رفع ابهام بدون نظارت/بی‌سرپرست^{۱۳}. زیرشاخه‌های این رویکرد می‌تواند شامل طبقه‌بندی بافتی^{۱۴}، طبقه‌بندی واژه^{۱۵} و گراف هم‌رخدادی^{۱۶} شود.

برخی از این رویکردها در زبان فارسی نیز به کار گرفته شده‌اند، اما جا دارد نرم‌افزارهایی با صحت بالا به‌منظور رفع ابهام از هم‌نگاره‌های فارسی ساخته شود که در این راستا اطلاعات به‌دست آمده از پیکره و دسته‌بندی انواع هم‌نگاره‌ها بسیار مفید خواهد بود.

1. knowledge-based WSD

2. LESK algorithm

3. semantic similarity

4. selectional preferences

5. heuristic method

6. supervised WSD

7. decision list

8. decision tree

9. Naïve Bayes

10. neural networks

11. instance-based learning

12. support vector machine

13. unsupervised WSD

14. context clustering

15. word clustering

16. co-occurrence graph

فهرست منابع

- خیرمند پاریزی، منیر، و رضا نورمندی‌پور. ۱۳۹۵. رفع ابهام معنایی کلمات فارسی با استفاده از روش یادگیری مبتنی بر نمونه. اولین کنفرانس بین‌المللی دستاوردهای نوین پژوهشی در مهندسی برق و کامپیوتر. تهران.
- دلجوان انوری، آزاده. ۱۳۹۷. رفع ابهام معنایی کلمات با استفاده از روش‌های مبتنی بر گراف. پایان‌نامه کارشناسی ارشد. دانشگاه تحصیلات تکمیلی علوم پایه. گاوازنگ. زنجان.
- شقاقی، ویدا. ۱۳۹۱. مبانی صرف. سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها (سمت). مرکز تحقیق و توسعه علوم انسانی.
- صادقی، علی‌اشرف ۱۳۷۱ الف. شیوه‌ها و امکانات واژه‌سازی در زبان فارسی معاصر (۵). نشر دانش: ۳۹-۴۵.
- _____. ۱۳۷۱ ب. شیوه‌ها و امکانات واژه‌سازی در زبان فارسی معاصر (۷). نشر دانش: ۷۲-۷۳.
- _____. ۱۳۷۱ ج. شیوه‌ها و امکانات واژه‌سازی در زبان فارسی معاصر (۸). نشر دانش: ۷۴-۹۸.
- _____. ۱۳۷۲ الف. شیوه‌ها و امکانات واژه‌سازی در زبان فارسی معاصر (۹). نشر دانش: ۷۵-۹۵.
- _____. ۱۳۷۲ ب. شیوه‌ها و امکانات واژه‌سازی در زبان فارسی معاصر (۱۰). نشر دانش: ۷۶-۹۵.
- علایی، الهام. ۱۳۹۷. معرفی رویکردی ماشینی با استفاده از الگوریتم لسک و برجسب‌دهی نحوی جهت رفع ابهام از معنی کلمات. پژوهشنامه پردازش و مدیریت اطلاعات ۳۳ (۳): ۱۱۹۶-۱۱۷۹.
- علی‌پوری حافظی، حامد، امیرسعید مولودی، و محمدکریم بیات. ۱۳۹۸. رفع ابهام معنایی از واژگان هم‌آوا-هم‌نویسه فارسی: رویکرد پیکره‌بنیاد. کنفرانس ملی بازیابی تعاملی اطلاعات. تهران.
- لازار، ژیلبر ۱۳۸۹. دستور زبان فارسی معاصر. ترجمه مهستی بحرینی و توضیحات و حواشی هرمز میلانیان. انتشارات هرمس. چاپ دوم.
- محسنی، مهدی. ۱۳۸۷. سیستم برجسب‌گذاری و ابهام‌زدایی خودکار اجزای کلام برای پیکره متنی زبان فارسی. تهران: دانشگاه علم و صنعت. دانشکده مهندسی کامپیوتر.
- مسعودی، بابک، و سعید راحتی قوچانی. ۱۳۹۴. رفع ابهام معنایی واژگان مبهم فارسی با مدل موضوعی LDA. فصلنامه پردازش علائم و داده‌ها. ۲۶ (۴): ۱۱۷-۱۲۵.
- مظفری، زهرا، گیتی تاکی، مجتبی صباغ جعفری، و پاکزاد یوسفیان. ۱۳۹۷. سامانه رفع ابهام معنایی از حروف اضافه در زبان فارسی با استفاده از قالب‌های معنایی. پژوهش‌های زبانی ۹ (۱): ۹۹-۱۱۷.

References

- Alqahtani, Sawsan, Hanan Aldarmaki, & Mona Diab. 2019. Homograph disambiguation through selective diacritic restoration. *Proceedings of the Fourth Arabic Natural Language Processing Workshop*. Florence, Italy. Association for Computational Linguistics. 49-59.
- Bakx, G. E. 2006. *Machine learning techniques for Word Sense Disambiguation*.: Universitat Politècnica de Catalunya. Barcelona.
- Bijankhan, M., J. Sheykhzadegan, M. Bahrani, & M. Ghayoomi. 2011. Lessons from Building a Persian Written Corpus: Peykare. *Language Resources and Evaluation* 45 (2): 143-164.

- Gaustad, T. 2004. *Linguistic Knowledge and Word Sense Disambiguation*. Netherland: University of Groningen.
- Merriam-Webster Dictionary. <https://www.merriam-webster.com/dictionary> (accessed March 4, 2022)
- Pal, Alok Ranjan., and Dganta Saha. 2015. Word sense disambiguation: a survey. *International journal of control theory and computer modeling (IJCTCM)* 5 (3): 1-16.
- Twilley, Leslie, C. Peter Dixon, Dean Taylor, & Karen Clark. 1994. University of Alberta norms of relative meaning frequency for 566 homographs. *Memory and cognition* 22: 111-126.
- Yarowsky, D. 1997. *Homograph disambiguation in text-to-speech synthesis*. *Progress in speech synthesis*.157-172. New York Springer.

پیوست ۱. هم‌نگارها و فراوانی برجسب‌ها

هم‌نگارها	برجسب ۱	برجسب ۲	برجسب ۳	برجسب ۴	برجسب ۵	برجسب ۶
بزاری	N:144	AJ:70				
اتفاقی	N:144	AJ:96	ADV:9			
احساساتی	AJ:30	N:12	ADV:2			
احساسی	N:85	AJ:85	ADV:3			
اختیاری	AJ:91	N:39	ADV:2			
ارتباطی	AJ:621	N:179				
ارتشی	AJ:40	N:29				
اسلامی	AJ:20350	N:102	ADV:1			
اسنادی	N:39	AJ:24				
اشتباهی	N:51	AJ:16	ADV:4			
اصطلاحی	N:30	AJ:14				
اصفهانی	N:117	AJ:48				
اصلی	AJ:3923	N:59				
اصولی	AJ:333	N:58				
اطلاعاتی	AJ:607	N:278				
اعتباری	N:200	AJ:200				
اعتقادی	AJ:187	N:60				
اعظمی	AJ:23	N:10				
افتخاری	N:53	AJ:52				
باستانی	AJ:457	N:28				
باشگاهی	AJ:32	N:10				
افزایشی	N:40	AJ:21				
افغانی	AJ:117	N:25				
افقی	AJ:119	N:47				
اقتصادی	AJ:7136	N:27	ADV:2			
اقلیمی	AJ:103	N:10				
انبوهی	N:31	AJ:21	DET:7			
انتخاباتی	AJ:305	N:58				
انتخابی	AJ:215	N:49				
انتظامی	AJ:1719	N:20				
انتقادی	AJ:125	N:33				
انحرافی	AJ:82	N:17				

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
باطنی	AJ:55	N:13				
باغی	AJ:47	N:32				
بافتی	AJ:24	N:20				
باقی	N:1431	AJ:53				
اندکی	AJ:195	DET:134	ADV:73			
بیری	N:14	V:9	AJ:3			
امنیتی	AJ:1408	N:14				
انسانی	AJ:2358	N:284	ADV:1			
انضباطی	AJ:120	N:25				
انعکاسی	N:16	AJ:13				
انفجاری	AJ:39	N:26				
انفرادی	AJ:152	N:11				
انقلابی	AJ:856	N:115	ADV:2			
ایالتی	AJ:139	N:14				
بازاری	AJ:57	N:31				
ایرانی	AJ:3117	N:177				
بحرانی	AJ:160	N:54				
ایزدی	N:30	AJ:16				
آزادی	N:2753	AJ:11	V:5			
بازرگانی	N:1154	AJ:94				
آزمایشی	AJ:243	N:44	ADV:1			
آسانی	N:48	AJ:39				
ایمانی	AJ:88	N:37				
باندی	AJ:27	N:14				
بانکی	AJ:702	N:15				
آوازی	AJ:18	N:16				
بادی	AJ:77	N:15	V:2	RES:1		
آوایی	AJ:85	N:10				
ایمنی	N:255	AJ:216				
آیینی	AJ:51	N:32				
آبادانی	N:116	AJ:12				
آبزی	AJ:21	N:11				
آرمانی	AJ:77	N:12				
آذری	N:54	AJ:40				

هم‌تکاه‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
باری	N:107	AJ:77	RES:34	CONJ:28		
آبی	AJ:518	N:124				
آماري	AJ:295	N:69				
بارانی	AJ:65	N:26				
آموزشی	AJ:2604	N:22				
آرامی	N:83	AJ:31				
بختیاری	N:106	AJ:10				
بدیعی	AJ:20	N:10				
برتری	N:513	AJ:13	V:1			
برنامه‌ای	N:265	AJ:29				
برقی	AJ:132	N:15				
برهانی	N:15	AJ:11	V:1			
بزرگتری	AJ:32	V:20				
بزرگواری	N:41	AJ:10	V:1			
بزرگی	AJ:928	N:177				
بدی	AJ:135	N:105	V:16			
بستری	AJ:150	N:94				
بسیاری	DET:2814	AJ:736	PRO:143			
بسیجی	AJ:156	N:89				
بشری	AJ:635	N:21				
بعدی	AJ:888	N:21	RES:3	PRO:1		
بلندی	N:88	AJ:78				
بندری	AJ:53	N:25				
بودجه‌ای	N:47	AJ:18				
بهاری	AJ:60	N:12				
بهشتی	N:367	AJ:24				
بی‌خبری	N:26	AJ:4	V:1			
بیرونی	AJ:190	N:29				
بیمارستانی	AJ:99	N:36				
پاره‌ای	DET:342	PRO:16				
پایانی	AJ:558	N:29				
پایداری	N:236	AJ:14				
پدری	N:88	AJ:26				
پرداختی	AJ:76	N:12				

هم‌نگارها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
پرسشی	N:69	AJ:24				
پروازی	AJ:59	N:10				
پروتئینی	AJ:75	N:10				
پزشکی	N:1891	AJ:222				
پژوهشی	AJ:624	N:64				
پلیسی	AJ:63	N:14				
پنهانی	AJ:71	ADV:11	N:1			
پوششی	N:31	AJ:16				
پولی	AJ:212	N:92				
پستی	AJ:346	N:55				
پهلوانی	AJ:35	N:13				
پهلوی	N:323	AJ:25	P:17			
پیش‌دانشگاهی	N:100	AJ:21				
پیشگامی	N:16	AJ:12				
پیشنهادی	AJ:297	N:81				
پیمانی	N:28	AJ:15				
تاریخی	AJ:2480	N:68				
تاریکی	N:181	AJ:16				
تبریزی	N:70	AJ:15				
تبلیغاتی	AJ:484	N:33	ADV:1			
تجهیزاتی	N:23	AJ:18				
تحقیقاتی	AJ:621	N:109				
تحقیقی	AJ:159	N:88				
تحلیلی	AJ:117	N:77				
تخصصی	AJ:848	N:11	ADV:1			
ترتیبی	N:106	AJ:23				
ترکی	AJ:133	N:52	V:1			
ترکیبی	N:139	AJ:87	ADV:1			
تسخیری	N:16	AJ:11				
تشکیلاتی	AJ:87	N:43				
تصویری	AJ:297	N:227				
تضمینی	AJ:52	N:37				
تعادلی	AJ:26	N:10				
تعالی	AJ:223	N:121	RES:10			

هم‌نگارها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
تعاونی	N:324	AJ:29				
تعهدی	N:50	AJ:16				
تفسیری	N:68	AJ:47				
تقلیدی	AJ:20	N:13				
تکلیفی	N:40	AJ:28				
تلخی	N:60	AJ:46				
تلفنی	AJ:411	ADV:33	N:6			
تلفیقی	N:31	AJ:17				
تندروی	N:20	AJ:12				
تنگستانی	N:14	AJ:10				
تنی	N:90	RES:28	AJ:4			
توافقی	N:86	AJ:17				
توانایی	N:668	AJ:12				
توانمندی	N:83	AJ:10				
توانی	N:19	V:11	AJ:1			
توجهی	AJ:48	N:35				
توحیدی	N:45	AJ:34				
توده‌ای	AJ:25	N:14				
توری	N:13	AJ:11				
توسعه‌ای	AJ:38	N:25				
توصیفی	AJ:96	N:15				
توضیحی	N:74	AJ:15				
تولیدی	AJ:978	N:40				
تهرانی	AJ:210	N:66				
تیموری	N:50	AJ:24				
تیمی	AJ:105	N:61	RES:10			
تناثری	AJ:41	N:20	ADV:1			
ثابتی	AJ:48	N:11				
جاده‌ای	AJ:93	N:25				
جاسوسی	N:397	AJ:57				
جانوری	AJ:37	N:17				
جدی	AJ:1755	ADV:39	RES:5	N:3		
جدیدی	AJ:1228	N:15				
جرقه‌ای	AJ:12	N:11				

هم‌نگارها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
جزیره‌ای	N:27	AJ:14	RES:2			
جزیی	AJ:84	N:52				
جزئی	AJ:105	N:86	ADV:1			
جسمی	AJ:244	N:24				
جعفری	N:262	AJ:18				
جلدی	RES:16	AJ:11	N:5			
جمعی	DET:693	AJ:587	N:6			
جمعیتی	AJ:143	N:109	DET:1			
جناحی	AJ:205	N:38	ADV:3			
جنبشی	N:17	AJ:12				
جنسی	AJ:213	N:24				
جنگی	AJ:493	N:121				
جوانی	N:524	AJ:70	V:2			
جوهری	N:27	AJ:17				
جوی	AJ:160	N:89	V:5			
جهانی	AJ:5275	N:125	V:1	ADV:1		
جهشی	N:19	AJ:12				
چشمی	N:54	AJ:54				
چینی	AJ:430	N:41				
حامی	N:121	AJ:39				
حرفه‌ای	AJ:726	N:15	ADV:2			
حرکتی	N:206	AJ:70				
حسابی	N:77	ADV:35	AJ:15			
حسی	AJ:116	N:28	ADV:1			
حسینی	N:346	AJ:88				
حضوری	N:122	AJ:119	ADV:2			
حقوقی	AJ:1015	N:100				
حقیقی	AJ:543	N:34				
حکمی	N:159	AJ:23				
حکومتی	AJ:388	N:122				
حمایتی	AJ:139	N:29				
حیاتی	AJ:347	N:39	ADV:2	RES:1		
حیوانی	AJ:123	N:42				
خارجی	AJ:4042	N:33				

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
خامی	N:12	AJ:12				
خانوادگی	AJ:345	N:18	ADV:1			
خبری	AJ:879	N:246	ADV:1			
خداوندی	AJ:85	N:23				
خدایی	N:75	AJ:63	V:5			
خدائی	N:10	AJ:10				
خدماتی	AJ:256	N:68				
خدمتی	N:43	AJ:21				
خراسانی	AJ:50	N:46				
خروجی	AJ:121	N:101				
خشکی	N:140	AJ:14				
خصوصی	AJ:1422	ADV:16				
خطی	AJ:231	N:77	RES:1			
خلقی	AJ:27	N:18				
خودی	AJ:144	N:49	V:2			
خوراکی	N:70	AJ:63				
خورشیدی	AJ:158	N:30				
خوبی	AJ:1199	N:85	V:3			
خوشی	N:75	AJ:51				
خونی	AJ:155	N:26				
خیابانی	AJ:135	N:37				
خیالی	AJ:57	N:21				
خیری	N:40	AJ:13				
داده‌ای	AJ:29	V:18	N:4			
دارویی	AJ:353	N:49				
دارویی	AJ:35	N:10				
داستانی	AJ:162	N:114				
دانایی	N:38	AJ:12				
دانشجویی	AJ:863	N:21	V:1			
دانشگاهی	AJ:643	N:53				
دانه‌ای	AJ:22	N:14				
درآمدی	N:61	AJ:46	V:1			
درختی	N:61	AJ:30				
درخشانی	AJ:64	N:42				

هم‌نگارها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
درخواستی	N:29	AJ:14				
درستی	AJ:152	N:95				
درسی	AJ:462	N:83				
درصدی	RES:135	N:39	AJ:7	NUM:2		
درمانی	AJ:839	N:24				
دریافتی	AJ:220	N:20				
دریایی	AJ:791	N:47				
دستگاهی	N:81	AJ:12				
دستوری	N:46	AJ:44				
دری	N:73	AJ:35				
دستی	AJ:452	N:99	ADV:3			
دشواری	AJ:94	N:69				
دفاعی	AJ:567	N:10				
دفتری	AJ:37	N:33				
دقیقه‌ای	RES:64	N:28				
دقیقی	AJ:117	N:34				
دمی	N:32	AJ:11				
دنایی	N:96	AJ:32				
دلاری	RES:33	AJ:10	N:9			
دورانی	N:75	AJ:22				
دوره‌ای	AJ:113	N:110				
دوری	N:229	AJ:15	CL:1			
دولتی	AJ:2133	N:170				
دیداری	N:157	AJ:43				
دهی	V:30	N:27				
دیدی	N:56	V:49				
دیکتاتوری	N:127	AJ:31				
دینی	AJ:2136	N:93				
دیواری	AJ:52	N:30				
ذهنی	AJ:425	N:19	ADV:1			
رابطه‌ای	N:112	AJ:12				
راحتی	N:101	AJ:28	V:1			
راستی	ADV:112	N:45	AJ:11			
راهنمایی	N:565	AJ:14				

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
ربانی	AJ:37	N:33	RES:1			
ربطی	N:75	AJ:26				
رحمانی	N:208	AJ:14				
رسانه‌ای	AJ:267	N:40				
رسمی	AJ:1919	N:11	ADV:3			
رفتاری	AJ:165	N:135				
رفیعی	N:32	AJ:13				
رقابتی	AJ:120	N:36				
رقمی	N:54	RES:13	AJ:4			
رمزی	AJ:33	N:16				
رنگی	AJ:201	N:57				
روانی	AJ:721	N:17				
روایی	AJ:38	N:17				
روحانی	N:317	AJ:271				
روحي	AJ:491	N:57	RES:2			
روزنامه‌ای	N:55	AJ:17				
روستایی	AJ:893	N:65				
روسی	AJ:442	N:10				
روشنی	AJ:184	N:102				
روی	N:3989	P:3865	V:2			
ریشه‌ای	AJ:54	N:12	ADV:2			
ریالی	AJ:157	RES:45	N:3			
زبانی	AJ:189	N:127	ADV:1			
زخمی	AJ:384	N:84	ADV:1			
زرتشتی	AJ:28	N:17				
زشتی	N:46	AJ:18				
زمانی	N:1973	AJ:514				
زمینی	AJ:837	N:170	ADV:1			
زندانی	AJ:230	N:148	ADV:1			
زیادی	AJ:3330	N:22	ADV:9			
زیبایی	N:276	AJ:140				
زیبائی	N:84	AJ:11	V:2			
ژاپنی	AJ:334	N:27				
ژنی	N:19	AJ:19				

هم‌نگارها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
ساختاری	AJ:242	N:42				
ساختمانی	AJ:351	N:48				
ساختی	AJ:16	V:9	N:4			
ساری	N:147	AJ:10				
سازمانی	AJ:260	N:122				
سازی	N:13	AJ:12	V:6			
ساعتی	N:92	AJ:22				
ساعی	N:43	AJ:10				
سامی	AJ:16	N:14				
سیکی	N:34	AJ:22				
سپاهی	N:37	AJ:21				
ستاره‌ای	N:23	AJ:19	RES:1			
سختی	N:260	AJ:164				
سردی	N:62	AJ:14				
سرزمینی	N:82	AJ:23				
سرعتی	N:48	AJ:22				
سرمایه‌ای	AJ:48	N:40				
سری	N:258	CL:251	AJ:120			
سریالی	N:69	AJ:10				
سطحی	AJ:244	N:92	ADV:5			
سفیدی	AJ:24	N:19				
سلجوقی	N:157	AJ:34				
سلطنتی	AJ:141	N:46				
سلیقه‌ای	AJ:41	N:12	ADV:1			
سنتی	AJ:1009	N:19	RES:3	ADV:1		
سنگی	AJ:212	N:38				
سنگینی	AJ:227	N:112				
سنی	AJ:688	N:102				
سواری	AJ:153	N:91				
سودمندی	N:18	AJ:18				
سوری	AJ:116	N:13				
سوزنی	AJ:44	N:10				
سوئی	AJ:12	N:11				
سیاهی	N:48	AJ:31				

هم‌نگاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
سیستمی	N:69	AJ:25				
سیگاری	AJ:80	N:20				
سینمایی	AJ:685	N:13	ADV:1			
شادابی	N:60	AJ:15				
شاکی	N:106	AJ:15				
شاهنشاهی	AJ:58	N:26				
شاهی	N:39	AJ:28	CL:3			
شبکه‌ای	N:61	AJ:14				
شجاعی	N:49	AJ:16				
شخصی	AJ:846	N:359				
شخصیتی	N:188	AJ:100				
شعاری	N:21	AJ:18				
شعری	N:92	AJ:71	RES:1	NUM:1		
شغلی	AJ:404	N:42				
شکلی	N:265	AJ:31				
شماره‌ای	RES:19	N:14	AJ:1			
شمسی	AJ:223	N:33				
شناختی	AJ:141	N:69				
شورایی	N:22	AJ:20				
شورشی	AJ:53	N:22				
شرطی	AJ:62	N:53				
شرقی	AJ:2067	N:40				
شنوایی	N:55	AJ:14				
شهابی	N:14	AJ:10				
شهرستانی	AJ:57	N:14				
شهروندی	N:46	AJ:25				
شهری	AJ:1287	N:248				
شیرازی	N:258	AJ:50	RES:1			
شیری	AJ:28	N:23				
شیرینی	N:164	AJ:11				
صادقی	N:195	V:17	AJ:10			
صبحگاهی	AJ:31	N:14				
صحرائی	AJ:40	N:11				
صحنه‌ای	N:55	AJ:18				

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
صدایی	N:117	AJ:14				
صدری	N:26	AJ:10				
صفاتی	N:28	AJ:14				
صفتی	AJ:16	N:15				
صفحه‌ای	RES:26	N:18	AJ:3			
صنعتی	AJ:2069	N:47				
صورتی	N:790	AJ:35				
صیادی	N:50	AJ:46				
ضربه‌ای	N:65	AJ:26				
ضروری	AJ:1135	ADV:19				
طبقاتی	AJ:137	N:16				
طلایی	AJ:207	N:30				
طوفانی	AJ:24	N:17				
طولی	AJ:36	N:13				
طیفی	N:25	AJ:13	DET:2			
ظاهری	AJ:228	N:22				
عاملی	N:181	AJ:36				
عامی	AJ:18	N:13				
عبادی	AJ:60	N:46	RES:4			
عباسی	N:173	AJ:13				
عددی	AJ:42	N:14	RES:2			
عده‌ای	N:739	DET:165				
عراقی	AJ:315	N:69				
عزیزی	N:112	AJ:24				
عشایری	AJ:104	N:12				
عشقی	N:46	AJ:23	V:1			
عضوی	N:62	RES:10	AJ:4			
عظیمی	AJ:309	N:32				
علمی	AJ:3930	N:50	ADV:5	RES:1		
عملکردی	N:20	AJ:17				
عملی	AJ:1152	N:160	ADV:6			
عملیاتی	AJ:271	N:75				
عهدی	N:15	AJ:11	RES:1			
عینی	AJ:358	N:19	RES:4			

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
غذاخوری	N:27	AJ:20				
غذایی	AJ:1222	N:31				
غذائی	AJ:86	N:11				
غربی	AJ:2294	N:35				
غلطی	AJ:37	N:14				
غنی	AJ:288	N:14	RES:5			
غیرخودی	AJ:23	N:12				
غیرنظامی	AJ:212	N:33				
فراری	AJ:109	N:20				
فرانسوی	AJ:601	N:14				
فارغ‌التحصیلی	AJ:23	N:17				
فراوانی	AJ:678	N:106				
فردی	AJ:618	N:539				
فرضی	AJ:77	N:11				
فرعی	AJ:282	N:12	ADV:1			
فرودگاهی	AJ:14	N:13				
فرهنگی	AJ:6985	N:84	ADV:1			
فشاری	N:48	AJ:24				
فصلی	AJ:69	N:66				
فضایی	AJ:363	N:249				
فضائی	AJ:77	N:24				
فعلی	AJ:869	N:26	ADV:1			
فقیری	AJ:14	N:11				
فکری	AJ:915	N:137				
فلسطینی	AJ:638	N:43				
فلسفی	AJ:486	N:18				
فنی	AJ:1576	N:15				
فولادی	AJ:73	N:18				
قاجاقچی	N:75	AJ:24				
قالی	N:34	AJ:20				
قانونی	AJ:1938	N:141	ADV:7			
قبولی	N:96	AJ:38	RES:1			
قدرتی	N:190	AJ:22				
قدری	N:273	ADV:38	DET:9	AJ:2		

هم‌نگارها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
قدسی	AJ:28	N:25				
قدیمی	AJ:801	N:16				
قراردادی	N:164	AJ:74				
قرآنی	AJ:778	N:13				
قربانی	N:307	AJ:39				
قرمزی	N:19	AJ:13				
قسمتی	N:470	DET:11	ADV:10	RES:3		
قشری	N:42	AJ:10				
قطبی	AJ:41	N:11				
قطره‌ای	AJ:28	N:13				
قلبی	AJ:676	N:26	RES:5			
قلمی	AJ:21	N:17				
قندی	AJ:65	N:18				
قومی	AJ:317	N:36	RES:3			
قهرمانی	N:1133	AJ:129	V:1			
قیمتی	N:87	AJ:34				
کارآمدی	N:65	AJ:22				
کاربردی	AJ:231	N:13				
کارشناسی	N:627	AJ:14				
کارکردی	AJ:28	N:24				
کارگاهی	AJ:25	N:14				
کارگری	AJ:195	N:34				
کاری	N:1507	AJ:504	V:2			
کاغذی	AJ:35	N:21				
کاهشی	N:16	AJ:10				
کتابخانه‌ای	AJ:40	N:18				
کتابی	N:305	AJ:15	RES:1			
کردی	V:163	AJ:63	N:27			
کرمانی	N:111	AJ:24				
کراهی	AJ:23	N:10				
کشوری	N:596	AJ:358	ADV:1			
کلاسی	N:19	AJ:17				
کلامی	AJ:230	N:23	ADV:1			
کلی	AJ:1505	N:73	ADV:26	DET:13		

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
کلیدی	AJ:226	N:16				
کمالی	N:72	AJ:14				
کمدی	N:55	AJ:51				
کمربندی	N:29	AJ:11				
کمکی	N:105	AJ:58				
کمی	AJ:590	ADV:308	DET:97	N:95		
کناری	N:55	AJ:54				
کنترلی	AJ:20	N:18				
کنی	V:417	N:35	AJ:1			
کو‌تاهی	AJ:265	N:101				
کو‌چکی	AJ:319	N:24				
کوفی	N:177	AJ:46				
کوهی	N:63	AJ:18				
گازی	AJ:160	N:11				
گر‌جی	N:39	AJ:10				
گر‌مایی	AJ:44	N:13				
گر‌می	N:95	AJ:17	RES:12			
گرو‌هی	N:951	AJ:450	ADV:1			
گزارشی	N:752	AJ:21				
گشتی	AJ:19	N:15	V:7			
گفتاری	AJ:68	N:23	ADV:1			
گلی	N:85	AJ:65				
گیاهی	AJ:263	N:37				
گیلانی	N:25	AJ:16				
لا‌جوردی	N:12	AJ:12				
لایه‌ای	AJ:45	N:19				
لحظه‌ای	N:201	AJ:12	ADV:1			
لشکری	AJ:32	N:20				
لفظی	AJ:65	N:15				
مادری	AJ:78	N:58				
مالیاتی	AJ:283	N:14				
ماند‌گاری	N:53	AJ:22				
ماهواره‌ای	AJ:364	N:14				
مبنایی	N:32	AJ:19				

هم‌نگاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
متری	RES:195	N:12	AJ:1			
مجاری	N:110	AJ:20				
محتوایی	AJ:41	N:10				
محلی	AJ:1449	N:217				
محمدی	N:284	AJ:59				
محوری	AJ:148	N:13				
محیطی	AJ:166	N:125				
مخزنی	AJ:27	N:10				
مداری	AJ:18	N:12				
مدرسه‌ای	N:30	AJ:10				
مدنی	AJ:594	N:47	RES:1			
مدیریتی	AJ:169	N:19				
مذهبی	AJ:1366	N:16				
مرحله‌ای	N:76	AJ:70	RES:2			
مردمی	AJ:1029	N:294	V:2			
مرده‌ای	N:11	AJ:10	V:1			
مرکزی	AJ:3628	N:126				
مسافربری	AJ:160	N:31				
مسافری	AJ:95	N:22	NUM:1			
مستمری	N:66	AJ:17				
مسجدی	N:39	AJ:14				
مسلمانی	N:87	AJ:22				
مسیحی	AJ:279	N:25				
مشارکتی	AJ:71	N:16				
مطالعاتی	AJ:112	N:44				
مطبوعاتی	AJ:651	N:24				
مصری	AJ:143	N:11				
معدودی	AJ:65	DET:21	N:8			
معدنی	AJ:389	N:13				
معرفتی	AJ:30	N:16	RES:1			
معروفی	AJ:38	N:11				
معماری	N:580	AJ:24				
معمایی	AJ:28	N:10				
معنایی	AJ:94	N:89				

هم‌نگارها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
معنی	AJ:163	N:12				
مفهومی	N:114	AJ:59				
مفیدی	AJ:69	N:10				
مقادیری	N:32	DET:28				
مقامی	N:82	AJ:38				
مقایسه‌ای	N:26	AJ:21				
مقداری	DET:167	N:165	ADV:33	AJ:8		
مقدماتی	AJ:475	N:15				
مقرراتی	N:45	AJ:12				
مقطعی	AJ:71	N:45	ADV:1			
مکانی	N:95	AJ:71				
مکتبی	N:22	AJ:17				
ملکی	N:61	AJ:30				
منطقه‌ای	AJ:986	N:187	ADV:3			
منطقی	AJ:587	N:18	ADV:7			
موتوری	AJ:42	N:13				
موجی	N:55	AJ:24				
موردی	N:98	AJ:42				
موشکی	AJ:214	N:16				
موضعی	N:77	AJ:46				
موضوعی	N:194	AJ:125				
مهارتی	N:14	AJ:10				
مهربانی	N:90	AJ:12	V:1			
مهندسی	N:542	AJ:235				
میدانی	AJ:34	N:32	V:11			
میلیونی	AJ:43	RES:27	NUM:2			
مینوی	AJ:16	N:15				
ناحیه‌ای	N:46	AJ:15				
نادرستی	N:32	AJ:27				
نادری	N:46	AJ:32				
ناراحتی	N:215	V:20				
ناراضی	AJ:69	N:18	ADV:1			
نازی	AJ:16	N:10				
نامی	N:149	AJ:137	NUM:1			

هم‌تکارها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
نبوی	N:131	AJ:45				
نحوی	N:552	AJ:178				
نزدیکی	N:675	AJ:70	V:1			
نژادی	AJ:163	N:11				
نشستی	N:129	V:11				
نظارتی	AJ:93	N:16				
نظامی	AJ:3437	N:275				
نظری	AJ:401	N:326				
نفتی	AJ:821	N:20				
نفری	N:79	RES:76	AJ:1			
نفوذی	AJ:50	N:35				
نقدی	AJ:346	N:63	ADV:2			
نقطه‌ای	N:114	AJ:11	RES:1			
نمادی	N:52	AJ:15				
نمایشگاهی	N:160	AJ:43				
نمایشی	AJ:320	N:53				
نمایی	N:37	V:10	AJ:2			
نمودی	N:15	V:14				
نمونه‌ای	N:219	AJ:15	DET:1			
نورانی	AJ:147	N:14				
نوروزی	AJ:220	N:71				
نوری	N:476	AJ:121				
نوشته‌ای	N:33	V:10				
نوعی	N:1514	AJ:18				
نهادی	N:78	AJ:65	V:5			
نیرومندی	AJ:40	N:12				
نیستی	V:74	N:37				
نیشابوری	N:45	AJ:13				
واحدی	AJ:99	N:58	RES:14			
واژگانی	AJ:257	N:17				
واسطه‌ای	AJ:20	N:11				
واکنشی	N:65	AJ:27				
وجودی	AJ:191	N:50				
ورزشی	AJ:1324	N:12				

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
ورودی	N:343	AJ:97				
وزنی	AJ:30	N:13				
وضعی	N:87	AJ:29				
وقتی	CONJ:2985	N:505				
ولایتی	N:36	AJ:32				
هادی	N:316	AJ:38				
هدایتی	N:17	AJ:11				
هستی	N:382	V:233				
هلالی	N:15	AJ:13				
همزمانی	N:32	AJ:11				
همگی	PRO:407	DET:35				
هندی	AJ:338	N:25				
هنری	AJ:1854	N:96	V:1	ADV:1		
هنگامی	N:768	CONJ:10				
هوایمایی	N:268	AJ:33				
هوایی	AJ:1046	N:25				
هویتی	AJ:24	N:20				
یاغی	N:30	AJ:16				
یکی	NUM:10055	AJ:49				
یونانی	AJ:285	N:11				
یهودی	AJ:285	N:82				
متجاوز	AJ:156	N:23				
متخصص	AJ:378	N:152				
عجم	AJ:21	N:16				
کافر	AJ:57	N:15				
متعلق	AJ:851	N:15				
متعلقه	AJ:14	N:11				
متغیر	N:163	AJ:136				
مانع	N:846	AJ:42				
مروج	N:24	AJ:16				
متفکر	N:59	AJ:30	ADV:1			
متقاضی	AJ:134	N:95				
متکلم	N:13	AJ:11				

هم‌نگارها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
متوجه	AJ:942	N:98				
متوسط	AJ:767	N:299				
متوسطه	AJ:200	N:45				
متوفی	AJ:50	N:24				
متولد	N:550	AJ:70				
مجاهد	AJ:41	N:19				
مجتهد	N:56	AJ:23				
اشتباه	N:573	AJ:82	ADV:11			
مبارز	AJ:322	N:21				
مبتکر	N:48	AJ:12				
متحد	AJ:1307	N:38	ADV:1			
مجرد	AJ:97	N:20				
متصل	AJ:341	ADV:10				
متقی	AJ:24	N:10				
متمم	AJ:33	N:30				
متهم	N:844	AJ:416	RES:1			
مجروح	AJ:291	N:73				
مجهول	AJ:29	N:14				
محافظ	AJ:61	N:28				
ضمن	P:2151	N:395				
ضمیمه	AJ:103	N:64				
طراح	N:118	AJ:10				
محبوب	AJ:166	N:35				
محتاج	AJ:124	N:16				
محبوب	N:38	AJ:16				
محرک	AJ:81	N:68				
مايع	N:195	AJ:68				
ظالم	AJ:65	N:26	RES:3			
ظاهر	N:743	AJ:23				
ظریف	AJ:125	N:26				
مايل	AJ:313	N:63	CL:9			
محصل	N:35	AJ:15				
محقق	N:377	AJ:214				
محکوم	AJ:1041	N:10				

هم‌تکاه‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
محلول	N:77	AJ:10				
مخالف	AJ:1032	N:58				
مختصر	AJ:165	N:17	ADV:2			
مخفف	N:20	AJ:10				
مخلوط	N:179	AJ:54	ADV:1			
مخلوق	N:22	AJ:19				
عاشق	AJ:180	N:66				
مدافع	AJ:169	N:68				
مداوم	AJ:296	ADV:10				
مراقب	N:95	AJ:71				
مرتب	AJ:158	ADV:92				
مستقل	AJ:1068	ADV:21	N:5			
مستلزم	N:267	AJ:66				
مستند	AJ:369	N:81	ADV:2			
مسئول	N:1313	AJ:350	RES:3			
مشابه	AJ:924	N:17				
مشاور	N:568	AJ:32				
مشترک	AJ:2211	N:54	RES:1			
مشروح	AJ:55	N:24	ADV:1			
مشمول	AJ:307	N:14				
مشوق	N:29	AJ:10				
مصدوم	AJ:27	N:12				
مصوب	AJ:412	N:19				
مطلع	AJ:199	N:140				
ضعیف	AJ:527	ADV:15	N:12	RES:2		
مطلق	AJ:503	N:14	RES:1			
علیه	P:2465	RES:64	N:2	AJ:1		
مطلوب	AJ:840	N:16	ADV:6			
مطهر	AJ:246	N:20				
بدیع	AJ:61	N:11				
مظلوم	AJ:164	N:23	RES:1			
مظنون	N:56	AJ:14				
معالج	AJ:59	N:30				
معتمد	N:18	AJ:17				

هم‌نگارها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
معرف	N:110	AJ:12				
حق	N:3472	AJ:25	RES:17			
معرق	N:28	AJ:26				
حامل	N:390	AJ:28	RES:1			
ثابت	AJ:1097	N:19	RES:4	ADV:1		
معروف	AJ:1126	N:159				
معصوم	AJ:102	N:44	RES:1			
معقول	AJ:243	N:22	ADV:1			
معلول	AJ:133	N:49				
خالص	AJ:229	N:14	ADV:1			
داخل	N:1634	P:614	AJ:9	RES:1	ADV:1	
کامل	AJ:2709	N:11	ADV:11			
رابط	N:61	AJ:24				
مفصل	AJ:107	N:12	ADV:7			
مقابل	N:2387	P:653	AJ:154	ADV:1		
منکر	N:265	AJ:30				
قادر	AJ:1091	N:20				
باطل	AJ:195	N:95	ADV:1			
دایم	AJ:64	ADV:20				
دائم	AJ:256	ADV:52				
حقیر	N:66	AJ:54				
فائق	AJ:45	N:16				
حاصل	N:1646	AJ:48				
حاکم	AJ:905	N:482				
خارج	N:3250	P:165	AJ:92			
ذیل	AJ:519	N:90	P:55			
حلال	AJ:85	N:17	RES:3			
دقیق	AJ:1080	ADV:25				
شریف	AJ:405	N:176				
فاتح	AJ:76	N:31				
فوق	AJ:1298	N:10	RES:5	ADV:1		
شریک	N:175	AJ:12	RES:3			
قانع	AJ:118	N:12				
قبل	P:2882	ADV:1188	RES:18			

هم‌نگاردها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
قبول	AJ:1001	N:446				
قبیل	N:703	CL:206				
قتل	N:1000	RES:15				
قدس	N:421	AJ:10	RES:2			
قائم	N:67	AJ:53	ADV:1			
سابق	AJ:1429	N:36	ADV:13	RES:2		
مقدس	AJ:1605	N:25	ADV:1			
مقصر	AJ:76	N:15	ADV:4			
رفیع	AJ:72	N:35				
رقیب	N:179	AJ:159				
مکرر	AJ:313	ADV:76				
مکمل	AJ:159	N:10				
مميز	N:53	AJ:17				
سید	N:2288	RES:12				
شاخص	N:501	AJ:138	NUM:1			
شاهد	N:1115	AJ:249				
جاهل	AJ:48	N:12	RES:1			
سهل	AJ:47	N:11				
تابع	N:259	AJ:39				
ناطق	N:181	AJ:18				
ناظر	AJ:326	N:72				
عامل	N:1525	AJ:16				
ناقض	AJ:28	N:13				
صادر	AJ:1150	N:112				
صریح	AJ:243	ADV:20	N:1			
صغیر	AJ:98	N:87				
فقیه	AJ:129	N:26				
فقیر	AJ:326	N:11	RES:4			
فقیه	N:348	AJ:151				
قریب	AJ:200	N:65	RES:5	DET:5		
منافق	AJ:31	N:10				
منتخب	AJ:240	N:215				
منتقد	AJ:60	N:48				
شهید	N:2481	AJ:150	RES:5			

هم‌تکارها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
شی	N:96	RES:10	V:2			
شیء	N:92	RES:26				
منتهی	AJ:322	CONJ:50	RES:1	CL:1		
جامع	AJ:867	N:19				
صبح	N:1586	ADV:22				
صحیح	AJ:1013	ADV:18	N:7			
منقول	AJ:35	N:17				
ضد	N:1142	AJ:12				
موافق	AJ:334	N:10				
موفق	AJ:1856	ADV:25	N:1			
عاقل	AJ:58	N:19	RES:1			
مولد	AJ:78	N:26	ADV:1			
سیال	AJ:111	N:32				
موحد	AJ:22	N:21				
معدود	AJ:103	CL:19				
مهاجر	N:124	AJ:94				
مهاجم	N:137	AJ:52				
مؤدب	N:21	AJ:12				
مؤسس	N:49	AJ:11				
موید	N:38	AJ:18	V:1			
وارد	AJ:272	N:95				
واقع	AJ:2404	N:258				
واقف	AJ:84	N:16				
آرد	N:143	V:11				
بیر	N:38	V:19				
پیا	N:12	AJ:10				
بازی	N:2722	AJ:13				
بدو	N:149	P:36	V:2			
بدان	P:575	V:62	N:8	RES:2		
برد	V:1226	N:87				
برآورد	N:487	V:23				
برآیند	V:51	N:46				
برده	V:706	N:34	AJ:4			

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
ولی	CONJ:6687	N:257	RES:4			
برداشت	N:504	V:233				
برنده	N:208	AJ:158	NUM:3	ADV:2		
بعد	P:3953	ADV:1605	CONJ:1441	N:482	RES:27	
به	P:266180	RES:45	AJ:24	N:16		
بهت	P:22	N:10				
بیان	N:3391	V:3	RES:1			
پاکت	N:105	CL:11	AJ:3			
پر	AJ:868	N:81	ADV:4			
پست	N:698	AJ:104				
پیدایش	N:373	AJ:10				
ترک	N:1014	AJ:188	RES:3			
تست	N:139	V:20	AJ:1			
بینی	N:142	V:16	RES:1			
جور	N:216	CL:95	AJ:26			
چکیده	N:50	AJ:24	V:19			
خرد	N:267	AJ:133				
خرم	AJ:50	N:30	V:2			
خرید	N:1528	V:72				
خلق	N:685	RES:10				
خم	AJ:98	N:26				
خورد	V:321	N:22	AJ:1			
خیر	AJ:243	RES:236	N:215	ADV:67		
درست	AJ:1251	ADV:677	N:14			
درگذشت	N:367	V:291				
درهم	N:45	AJ:27	P:9	CL:2	ADV:1	
جدا	AJ:892	ADV:15	N:1			
روم	N:232	V:10				
دم	N:261	P:56				
دور	N:1436	AJ:1285	P:226	CL:7	ADV:5	
دوره	N:4536	CL:55				
دوم	NUM:5322	ADV:303	RES:1			
دوی	NUM:62	N:18				
ده	NUM:1370	N:97	V:43			

هم‌تکارها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
دهم	NUM:307	V:131				
تو	PRO:3497	P:145	N:22	ADV:5		
دیر	ADV:217	N:33				
راند	V:53	N:14				
رب	N:43	RES:28				
رفت	V:1468	N:150	AJ:1			
رو	N:1526	POSTP:657	P:12	V:4		
گویا	ADV:227	AJ:33	N:6			
رود	N:423	V:169				
روژه	N:306	RES:78	AJ:20	V:2		
رویش	N:118	P:28				
زند	V:80	N:80				
زنم	N:12	V:10				
ساکت	AJ:131	ADV:11	N:4			
سازند	V:397	N:22				
سیک	N:430	AJ:234	ADV:2			
سیبل	N:51	RES:17				
سپرده	N:253	V:181	AJ:2			
سالم	AJ:866	ADV:32	N:23			
سر	N:4940	P:123	AJ:3	RES:2	CL:2	
سنجیده	AJ:68	V:23	ADV:3			
سنه	N:14	RES:10				
سواره	AJ:24	N:12	ADV:2			
سوخته	AJ:86	V:17	N:3			
سیر	N:1576	AJ:42	ADV:3	CL:2		
شست	N:15	V:13				
شکست	N:1747	V:66				
شکسته	V:194	AJ:95	N:15			
شور	N:351	AJ:67				
شوم	V:189	AJ:110	ADV:1			
صرف	N:1033	AJ:167	ADV:1			
عالم	N:1461	AJ:41	CL:2			
عمده	AJ:1025	DET:49				
عموم	DET:311	N:286	AJ:2			

هم‌تکاه‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
غریبه	AJ:33	N:16				
قد	N:127	RES:28	AJ:1			
بین	P:5177	N:2427	RES:27	V:19		
قلم	N:685	CL:63	RES:1			
قوی	AJ:815	ADV:14	N:7	RES:1		
قل	RES:25	N:12				
کل	N:2561	DET:746	AJ:127	RES:60		
کرد	V:33453	AJ:73	N:55			
کر	N:34	AJ:32	RES:1			
کشت	N:550	V:82				
کلاسه	N:32	RES:18				
کلیه	DET:1519	N:140	AJ:6			
بدم	V:17	AJ:17				
بر	P:36951	N:661				
کن	V:647	N:147	RES:5			
کند	V:11886	AJ:140	ADV:13	N:4		
کنه	V:61	N:12				
کنار	ADV:2205	P:472	N:4			
کی	ADV:301	N:21	RES:2	CONJ:1		
که	CONJ:186345	N:172	ADV:34	V:1	RES:1	
گرد	N:456	AJ:31				
تنگ	AJ:165	N:35	ADV:4			
گردان	N:116	V:14	AJ:5			
گرفته	V:5748	AJ:30	ADV:1			
در	P:315826	N:951	AJ:1			
له	RES:54	AJ:40	P:7	N:1		
گرم	AJ:800	CL:251	N:23	ADV:4		
گزیده	N:21	V:19	AJ:5	ADV:3		
لخته	AJ:30	N:18				
ماشینی	AJ:79	N:14				
مثل	ADV:2736	N:114	RES:7			
هل	N:32	RES:11				
مرد	N:1841	V:88	AJ:48			
مردم	N:16124	V:13				

هم‌تکارها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
مردہ	AJ:170	V:65	N:30	ADV:1		
مرکب	AJ:193	N:81				
مسکن	N:910	AJ:13	RES:1			
مشتی	N:22	DET:12	AJ:3			
مشرف	N:383	AJ:38				
مشت	N:106	CL:19				
من	PRO:12599	RES:550	CL:21	N:15		
مقدم	AJ:174	N:75				
منه	V:19	RES:16	N:1			
بری	N:28	V:17	AJ:11			
نزد	P:754	N:164	V:32			
نبرد	N:385	V:40				
نبود	V:1939	N:364				
نشست	N:1770	V:266	NUM:2	AJ:1		
نشسته	V:430	AJ:50	N:1			
نهی	N:205	V:34	RES:4			
نی	N:45	ADV:10	RES:1			
نه	ADV:4945	NUM:174	N:54	RES:46	V:4	CONJ:3
اشرف	N:69	AJ:33				
آزاد	AJ:2264	N:325	ADV:18			
احد	N:44	RES:16	AJ:1			
اعظم	AJ:240	N:24	RES:3			
اکرم	AJ:337	N:16				
اکبر	N:291	AJ:21	RES:2			
آدم	N:847	RES:12	AJ:3			
آزاده	AJ:52	N:49				
آرام	AJ:511	N:142	ADV:131			
جاوید	AJ:44	N:40				
حافظ	N:727	AJ:31	RES:3			
خجسته	AJ:49	N:10				
سعید	N:527	AJ:156				
سیما	N:611	RES:10				
سیمین	N:24	AJ:20				

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
شکیبا	AJ:18	N:10				
فرزانه	AJ:69	N:20				
سلیم	N:101	AJ:31	RES:2			
جبار	AJ:21	N:17				
رشید	AJ:69	N:44				
رئوف	N:15	AJ:11	RES:4			
زاهد	N:51	AJ:16				
شیوا	AJ:16	N:13	ADV:1			
صدیق	AJ:26	N:23				
صادق	AJ:249	N:212				
جابر	N:18	AJ:13				
صالح	AJ:203	N:108	ADV:2			
طیبه	AJ:42	N:29	RES:2			
شیرین	AJ:327	N:98	ADV:4			
سوزان	AJ:39	N:14				
جمیل	N:21	AJ:16	RES:2			
عابد	N:16	AJ:14				
عادل	N:95	AJ:59				
پیروز	AJ:412	N:14				
عارف	N:178	AJ:25				
عزیز	AJ:940	N:157	RES:4	ADV:1		
علی	N:3157	RES:280	AJ:10			
صفر	NUM:871	N:93				
فاضل	N:72	AJ:55				
کبری	N:47	AJ:10				
کبیر	AJ:213	N:25				
کریم	AJ:511	N:276	RES:6			
کریمه	AJ:43	N:30				
فروتن	AJ:22	N:12				
لطیف	AJ:84	N:11				
متین	N:60	AJ:51	ADV:3	RES:1		
مجید	N:396	AJ:154				
مختار	AJ:77	N:58	ADV:1			
مسلم	AJ:346	N:49	ADV:7	RES:2		

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
مسعود	N:402	AJ:11				
معین	AJ:393	N:126				
نازنین	AJ:17	N:13				
امین	N:160	AJ:67	RES:1			
هویدا	AJ:21	N:15				
یگانه	AJ:86	N:18	CL:17			
ایستاده	V:244	AJ:42	ADV:6	N:1		
انباشته	AJ:71	V:20	ADV:1			
باقی مانده	AJ:84	N:29	V:1			
باز مانده	AJ:46	N:22	V:10			
مانده	V:879	N:111	AJ:23			
آراسته	AJ:53	V:12	N:9			
آفریده	V:70	N:12	AJ:1			
بسته	V:543	AJ:246	CONJ:109	N:38	CL:13	ADV:1
بازنشسته	AJ:146	N:83	V:8			
برخاسته	AJ:132	V:55	N:2			
برآمده	AJ:81	V:31	N:1			
برگرفته	AJ:44	V:37	N:11			
برگزیده	AJ:319	V:248	N:58			
بیمه شده	AJ:23	N:12				
پرداخته	V:585	AJ:14	N:6			
پسندیده	AJ:53	V:24	N:3			
پوشیده	V:183	AJ:73	ADV:2			
پیچیده	AJ:377	V:27	ADV:2	N:1		
نادیده	AJ:408	V:10	N:1	ADV:1		
پیوسته	AJ:229	ADV:164	V:68			
تحصیل کرده	AJ:92	N:10				
تراشیده	V:17	AJ:10				
ترسیده	AJ:17	V:13				
خوابیده	V:29	AJ:20	ADV:1			
داده	V:6445	N:85				
خواسته	V:433	N:218	ADV:25			
خورده	V:290	AJ:18	N:6			

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
رسیده	V:1947	AJ:218				
زاده	N:41	V:25	AJ:3			
ساخته	V:1957	N:215	AJ:17			
شناخته	V:816	AJ:10				
فرستاده	V:134	AJ:117	N:100			
فرموده	V:228	N:25				
پخته	AJ:103	V:27	ADV:2			
فهمیده	V:67	N:19	AJ:3			
کشته	V:1162	N:239	AJ:33			
کشیده	V:630	N:14	AJ:13			
گذشته	ADV:5778	N:1235	V:337	AJ:1		
گسترده	AJ:1063	V:25				
گشوده	V:164	AJ:12				
گفته	V:2248	N:2130				
خوانده	V:425	N:37	AJ:8			
نر سیده	V:280	AJ:23				
گمشده	AJ:64	N:15	V:7			
بریده	AJ:55	V:22	N:13			
دیده	V:1389	N:126				
نامبرده	N:154	AJ:12				
نوشته	V:1151	N:988	AJ:7			
نهفته	AJ:166	V:41	ADV:1			
ناخواسته	AJ:114	ADV:47				
ورزیده	AJ:43	V:22	N:1			
یافته	V:2400	N:18				
افزوده	V:508	AJ:102	N:1			
اینچ	N:50	CL:17	RES:2			
بشکه	N:343	CL:142				
پیمانه	CL:19	N:15				
تخته	N:58	CL:22				
نفر	N:7120	CL:313				
دست	N:10856	CL:30				
بار	N:4716	CL:14				

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
تکه	N:41	CL:39				
جفت	N:59	CL:41				
دانگ	N:20	CL:13				
چه	CL:52	N:23	RES:1	AJ:1		
حلقه	N:223	CL:86				
درجه	N:1553	CL:41	RES:1			
دستگاه	N:1972	CL:844				
درصد	N:7566	CL:198				
دسته	N:1128	CL:24				
راس	N:259	CL:34	RES:1			
هکتار	N:555	CL:178				
پوند	N:171	CL:20				
رأس	N:386	CL:37	RES:1			
ردیف	N:357	CL:13	AJ:1			
ریشتر	N:52	CL:28				
رشته	N:1414	CL:67	V:1			
سانتی متر	N:110	CL:21				
سانتیگراد	AJ:140	CL:25				
سانتیمتر	N:69	CL:40				
عدد	N:418	CL:163				
عضو	N:2477	AJ:83				
تار	N:96	AJ:31	ADV:2	CL:1		
فروند	CL:357	N:45				
فقره	CL:212	N:43				
قبضه	CL:115	N:13				
قطعه	N:284	CL:238				
کیلو	CL:144	N:114				
کیلوگرم	N:808	CL:513				
کیلومتر	N:770	CL:161				
کیلومتر مربع	N:178	CL:23				
کیلووات	N:35	CL:35				
مگاوات	N:65	CL:17				
تا	P:16672	CONJ:9834	CL:324	N:37		
کیلوولت	N:50	RES:30	CL:2			

هم‌تکاه‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
لیتر	CL:116	N:113				
جلد	N:436	CL:165	AJ:1			
متر	N:930	CL:231				
متر مربع	N:449	CL:162				
متر مکعب	N:217	CL:133				
مجموعه	N:1910	CL:27				
مورد	N:17718	CL:156				
میلی گرم	N:50	CL:39				
میلی لیتر	CL:16	N:11				
میلی متر	N:58	CL:10				
میلیمتر	N:186	CL:27				
واحد	N:3872	AJ:356	CL:215	RES:4		
ایثارگر	AJ:26	N:24				
اشغالگر	AJ:80	N:10				
اندیشمند	N:39	AJ:26				
آفریننده	AJ:24	N:15				
آلاینده	AJ:43	N:25				
بازیگر	N:711	AJ:13				
بستانکار	N:55	AJ:12				
نیکوکار	AJ:81	N:11				
درستکار	N:12	AJ:12				
بدهکار	AJ:90	N:24				
ورزشکار	N:133	AJ:16				
درخشنده	N:15	AJ:14				
برگزارکننده	AJ:15	N:11				
بیانگر	N:318	AJ:65				
بازدارنده	AJ:86	N:14	V:7			
تأمین کننده	N:17	AJ:10				
تعیین کننده	AJ:199	N:10				
تقویت کننده	N:22	AJ:20				
توزیع کننده	N:14	AJ:13				
جنگنده	N:57	AJ:24				
تولید کننده	AJ:145	N:138				

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
جلوه گر	N:38	AJ:10				
خدمتگزار	N:66	AJ:35				
خر دمند	AJ:40	N:20				
هنرمند	N:348	AJ:48				
نیرومند	AJ:136	N:10	ADV:2			
دانشمند	N:236	AJ:40				
دریافت کننده	N:21	AJ:10				
شرکت کننده	AJ:307	N:76				
صادر کننده	N:95	AJ:61				
آینده	N:2771	AJ:1516				
سازنده	AJ:406	N:51	ADV:1			
رزمنده	N:28	AJ:14				
کاوشگر	N:23	AJ:15				
کارگر	N:586	AJ:20				
کارمند	N:227	AJ:18				
برومند	N:40	AJ:27				
کنترل کننده	AJ:33	N:14				
گیرنده	N:115	AJ:21				
مذاکره کننده	AJ:31	N:10				
مصرف کننده	N:118	AJ:34				
منعکس کننده	N:42	AJ:17				
نشان دهنده	AJ:146	N:96				
نشانگر	N:230	AJ:28				
هماهنگ کننده	N:59	AJ:15				
باد	N:637	V:158				
بازداشت	N:804	V:16	AJ:2			
بازگشت	N:892	V:167				
باز یافت	N:122	V:10				
برخاست	V:95	N:25				
برگشت	N:288	V:131				
بست	V:168	N:21				
پرداخت	N:2074	V:812				
پیوست	V:268	N:49	AJ:5			

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
خواست	V:1098	N:424				
داد	V:8016	N:200				
درآمد	N:1233	V:424				
داشت	V:9437	N:12				
دوخت	V:34	N:17				
دید	N:718	V:680				
رسید	V:3243	N:174				
روند	N:1954	V:53	AJ:2			
دان	N:28	V:11				
دریافت	N:3166	V:90				
زن	N:2240	AJ:92	V:15			
زاید	AJ:36	N:27	ADV:2			
زیست	N:1118	V:10				
ساخت	N:2290	V:1028				
ساز	N:123	V:11				
ست	V:337	N:55				
سرود	N:148	V:18				
سوخت	N:549	V:78				
شناخت	N:1201	V:125				
کاست	N:47	V:29				
گذشت	N:523	V:456				
گیر	N:72	V:11	AJ:7			
نواخت	N:167	V:10				
نوش	N:15	AJ:13	V:1			
نهاد	N:747	V:202				
یافت	V:3468	N:212	PUNC:1			
گشت	V:305	N:59				
نمود	V:1592	N:100				
آگاهانه	AJ:129	ADV:40				
آزادانه	ADV:96	AJ:33				
بیرحمانه	AJ:11	ADV:11				
جداگانه	AJ:325	ADV:39				
خالصانه	AJ:20	ADV:11				

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
خرمندان	AJ:69	ADV:10				
داوطلبانه	AJ:53	ADV:31				
دلسوزانه	AJ:33	ADV:13				
دوسالانه	N:39	AJ:29				
دوستانه	AJ:271	ADV:11				
دیوانه	AJ:40	N:33				
روزانه	ADV:687	N:18				
سالانه	AJ:172	ADV:109				
سرانه	AJ:179	N:72				
شجاعانه	AJ:27	ADV:25				
صادقانه	AJ:73	ADV:65				
صمیمانه	ADV:167	AJ:79	N:1			
عادلانه	AJ:251	ADV:20				
عاشقانه	AJ:80	ADV:21				
عجولانه	AJ:28	ADV:10				
فعالانه	ADV:29	AJ:28				
قاطعانه	ADV:53	AJ:24				
کورکورانه	AJ:14	ADV:10				
ماهانه	AJ:71	ADV:51	N:1			
مجدانه	AJ:13	ADV:10				
محرمانه	AJ:245	ADV:29				
مخفیانه	ADV:21	AJ:17				
مخلصانه	AJ:34	ADV:12				
مردانه	AJ:57	ADV:14				
میانه	AJ:257	N:108				
ناآگاهانه	AJ:18	ADV:15				
هوشمندانه	AJ:44	ADV:10				
هوشیارانه	ADV:14	AJ:12				
اول	NUM:5972	ADV:606	RES:3	N:2		
نخست	NUM:800	ADV:263				
آخر	NUM:909	ADV:215	RES:4			
پنجم	NUM:811	ADV:10				
چهارم	NUM:1118	ADV:18				

هم‌نگاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
یکباره	ADV:76	AJ:17				
دوباره	ADV:747	AJ:165				
یکپارچه	AJ:119	ADV:17				
یکتا	AJ:31	N:16	CL:1	ADV:1		
یکجا	ADV:47	AJ:20				
یکسان	AJ:338	ADV:17				
یکسره	ADV:47	AJ:29				
دو	NUM:18007	N:306				
سوم	NUM:3714	ADV:69				
سه	NUM:5714	N:23				
سه‌ماهه	N:57	AJ:44	ADV:2			
سی	NUM:662	N:44				
شش‌ماهه	N:29	AJ:29				
یک	NUM:45473	N:739				
تاجیک	AJ:69	N:61				
ترکمن	N:57	AJ:13				
صرب	N:38	AJ:12				
صفوی	N:156	AJ:137				
صفویه	N:200	AJ:15				
صهیونیست	AJ:83	N:21				
ضد انقلاب	N:67	AJ:24				
تازی	AJ:51	N:16				
تروریست	N:59	AJ:17				
جانباز	N:152	AJ:28	CL:1			
دموکرات	AJ:51	N:50				
سوسیالیست	AJ:76	N:20				
شیعه	N:501	AJ:42	RES:1			
غیرمسلمان	AJ:40	N:24				
فرانسه	N:4158	AJ:28				
سیاستمدار	N:95	AJ:13				
سیاه‌پوست	AJ:63	N:10				
علوی	N:72	AJ:60				
فیلسوف	N:147	AJ:16				

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
قاجار	N:299	AJ:21				
قزاق	N:23	AJ:17				
روس	N:147	AJ:78	RES:1			
کمونیسټ	AJ:331	N:29				
لاتین	N:135	AJ:76				
لیبرال	AJ:182	N:18				
لیبرال‌دمکرات	N:21	AJ:21				
مسلمان	AJ:985	N:228	ADV:1			
مومن	AJ:158	N:23	RES:1			
مؤمن	AJ:153	N:43	RES:3			
دیالکتیک	N:12	AJ:11				
هرمنوټیک	N:70	AJ:24				
اغلب	DET:578	ADV:519	PRO:27			
این‌جوری	ADV:19	AJ:11				
این‌چنین	ADV:55	PRO:18	AJ:9			
همین	DET:6895	PRO:540				
چون	CONJ:3729	ADV:1256	N:6			
چه	DET:4583	CONJ:1468	ADV:1270	N:886		
چنان	ADV:376	CL:334	PRO:239			
چند	DET:5464	AJ:43	ADV:7			
چندان	ADV:563	AJ:32				
چنین	CL:4128	PRO:1475	ADV:46			
کجا	ADV:560	N:55				
کدام	N:1099	DET:588	ADV:49			
چو	CONJ:134	ADV:85	N:6			
هم‌اکنون	ADV:1156	N:19				
هم‌زمان	AJ:54	ADV:15				
همان	DET:5301	PRO:217	RES:1			
همانند	ADV:617	AJ:14				
همچین	ADV:16	CL:11	AJ:1			
درحقیقت	CONJ:330	ADV:16				
بہتر	AJ:1593	ADV:238				
بعدازظهر	N:624	ADV:48				

هم‌تکاه‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
خلاصه	AJ:309	N:196	ADV:96			
بعض	DET:35	RES:10				
آهسته	ADV:88	AJ:31				
کم	AJ:1260	ADV:77	N:13	RES:10		
کمتر	AJ:1668	ADV:478				
صددرصد	ADV:131	DET:29	AJ:20	N:1		
بارها	ADV:611	N:37				
مادام‌العمر	AJ:23	N:11	ADV:3			
همه	DET:6810	PRO:2563				
خیلی	ADV:2720	DET:399	N:3	AJ:3		
هیچ	DET:6852	ADV:144	N:25	RES:9		
هیچی	PRO:29	RES:18	V:3	DET:2		
چی	ADV:297	N:51	DET:11			
سریع	AJ:581	ADV:75				
سریع‌تر	AJ:71	ADV:26				
سریعتر	AJ:180	ADV:61				
بلندمدت	AJ:183	ADV:58				
درازمدت	AJ:173	ADV:79				
این‌همه	DET:13	ADV:11				
به‌موقع	AJ:176	ADV:97				
آسان	AJ:295	ADV:17				
عاقبت	ADV:71	N:30				
مدام	ADV:93	AJ:14				
زود	ADV:178	AJ:138				
زیاد	AJ:1698	ADV:199	N:40	RES:2		
فوق‌العاده	AJ:455	ADV:102	N:11	DET:1		
زیادتر	AJ:71	ADV:12				
حتی	ADV:4855	RES:28	CONJ:1			
کوتاه‌مدت	AJ:245	ADV:75				
میان‌مدت	AJ:35	N:11				
حداقل	ADV:1083	N:286	DET:119	CL:66	AJ:15	
حداکثر	ADV:775	DET:253	CL:81	N:42	AJ:7	
حدود	DET:3868	N:505	RES:1			
بالقوه	AJ:281	ADV:25				

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
همزمان	AJ:1019	ADV:86				
بی‌نهایت	AJ:35	ADV:25	N:21			
بی‌وقفه	AJ:48	ADV:12				
سراسر	N:1206	DET:267	ADV:2			
سرانجام	ADV:632	N:44				
اینجا	ADV:1255	N:79				
فراسوی	N:23	P:16				
فراوان	AJ:838	ADV:18	N:1			
بحق	ADV:21	AJ:17	RES:1			
بحمدالله	ADV:169	RES:16				
اینگونه	ADV:129	DET:18				
تمام	DET:3852	AJ:895	RES:2	N:2		
تمام‌وقت	AJ:39	ADV:11				
پایین	AJ:450	ADV:325	P:39			
پائین	AJ:332	ADV:114	P:5			
تند	AJ:212	ADV:16	N:2			
تنها	ADV:3303	AJ:827	CL:652	N:3		
بمانند	V:144	ADV:11				
بموقع	AJ:43	ADV:11				
آنجا	ADV:1494	N:315				
برخی	DET:4513	PRO:421				
بسی	ADV:82	CL:19	DET:2			
فراتر	AJ:313	ADV:86				
بسیار	ADV:6584	AJ:490	DET:26			
بس	ADV:145	AJ:113	N:16			
بیش‌تر	ADV:67	DET:14	AJ:14			
بیشتر	AJ:5153	ADV:1592	DET:802			
بعضی	DET:2452	PRO:227				
اهل	N:1266	RES:23				
اهم	AJ:161	N:11				
ای	INT:726	N:38	DET:6	RES:3	PRO:1	
این	DET:138175	PRO:18333	RES:2	N:1		
اسیر	N:180	AJ:30	RES:2			
اضافه	AJ:1911	N:54				

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
اطراف	N:439	P:398	AJ:103			
اعلام	N:9693	AJ:28				
اعلم	RES:14	N:12	AJ:4			
اکثریت	N:555	DET:268				
انتخابیه	AJ:326	N:27				
آری	RES:167	N:19	V:2			
آن	PRO:41177	DET:15306	N:311			
آواره	AJ:85	N:28				
باب	N:644	CL:76	RES:4			
بابت	P:243	N:113				
باز	AJ:1620	ADV:738	N:8			
بازنده	N:17	AJ:16				
بالا	AJ:1965	ADV:636				
بالای	P:467	AJ:363	N:183			
بایر	AJ:24	N:10				
برابر	N:3461	P:381	AJ:241	ADV:1		
برات	N:20	P:15				
برعکس	ADV:190	AJ:68				
برقرار	AJ:786	N:36				
برگ	N:552	CL:19				
برنز	N:140	AJ:50				
برون	N:39	ADV:17	P:6	V:1		
بزرگ	AJ:5568	N:72	ADV:5			
بزرگسال	N:46	AJ:23				
بزرگوار	AJ:364	N:203				
بعید	AJ:155	ADV:12				
بل	CONJ:15	N:11	RES:10			
بلند	AJ:1106	ADV:19	N:16			
بله	RES:585	N:70	ADV:1			
بن	N:1085	RES:44				
بنی	N:27	RES:12				
بها	N:82	RES:12				
بهر	P:38	N:14				
بی‌تفاوت	AJ:82	ADV:10				

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
بی جهت	ADV:19	AJ:15				
بی سواد	AJ:113	N:18				
بی نظیر	AJ:182	N:21	ADV:1			
بیت	N:424	CL:13	RES:4			
بیچاره	AJ:51	N:40				
بیرون	N:1058	ADV:629	P:74	AJ:24		
بیکار	AJ:148	N:12				
بیگانه	AJ:304	N:71				
بیمار	N:869	AJ:108				
بیوتکنولوژی	N:67	AJ:18				
بیوه	N:11	AJ:10				
بیهوده	AJ:105	ADV:15				
پاره	AJ:88	N:16				
پایه	N:1030	AJ:171				
پس	P:8889	CONJ:1560	ADV:241	N:204	AJ:2	
پسر	N:914	AJ:41				
پشت	P:851	N:700	AJ:1			
پی‌درپی	AJ:111	ADV:14				
پیاده	AJ:483	ADV:20	N:10			
پیر	N:146	AJ:126				
پیرامون	P:788	N:65				
پیش	ADV:6071	P:571				
پیش‌کسوت	AJ:22	N:12				
پیشقدم	AJ:21	N:15				
تابناک	AJ:58	N:11				
تازه	AJ:815	ADV:290				
تام	AJ:78	N:24	CL:1			
تراژدی	N:47	AJ:26				
ترانزیت	N:104	AJ:95				
تعداد	N:3324	CL:147				
تعطیل	AJ:371	N:117				
تک	AJ:39	NUM:32	N:25	RES:1		
توانا	AJ:87	N:44				
توی	P:647	N:44				

هم‌نگاردها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
تیرانداز	AJ:13	N:11				
تیره	AJ:150	N:20				
تئوری	N:178	AJ:30				
ثمن	N:22	NUM:10	RES:2			
جان	N:2045	AJ:13				
جلو	ADV:376	P:140	AJ:71	N:1		
جلوتر	AJ:106	ADV:10				
جلوی	P:350	ADV:53				
جمع	N:1851	AJ:16	RES:2			
جمله	N:2654	CL:902	DET:29	PRO:7		
جنابعالی	PRO:585	N:35				
جوان	AJ:1493	N:626				
جهت	N:3021	P:2648				
حال	N:6357	ADV:472	RES:4			
حساب	N:1692	AJ:16				
حق‌شناس	N:19	AJ:14				
خاص	AJ:2349	N:68				
خاصه	ADV:59	AJ:18	N:11	RES:1		
خداحافظ	N:35	RES:13				
خلاف	N:1056	AJ:67	RES:1			
خلوت	N:71	AJ:39				
خواهان	N:557	AJ:95				
خوب	AJ:1913	ADV:634	N:19	RES:9		
خود	PRO:38785	N:21				
خودرو	N:1131	AJ:10	ADV:3			
خودگردان	AJ:276	N:25				
خوش	AJ:299	ADV:24	N:23			
دار	N:126	V:29	RES:4			
داغ	AJ:194	N:17				
دال	AJ:65	N:17	RES:1			
دانش‌پژوه	N:24	AJ:17				
دانشجو	N:755	AJ:46				
دانشجوی	N:233	AJ:10				
داوطلب	N:219	AJ:38	ADV:1			

هم‌نگارها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
دروغ	AJ:245	N:49				
دروغگو	N:12	AJ:11				
درون	N:940	P:249	AJ:1			
دریغ	ADV:109	N:53				
دست‌اندرکار	N:37	AJ:21				
دست‌کم	ADV:372	DET:11	AJ:8			
دستگیر	AJ:834	N:10				
دشمن	N:1735	AJ:17				
دقیقت	AJ:103	ADV:11				
دگر	PRO:43	DET:19	ADV:13			
دلاور	AJ:35	N:11				
دلخواه	AJ:112	N:17				
دنبال	N:2618	P:23				
دوچندان	ADV:27	AJ:21				
دوردست	AJ:82	N:11				
دوست	N:1297	AJ:18	V:2			
دیپلمه	N:18	AJ:12				
دیگر	PRO:14195	ADV:1579				
دیگه	ADV:121	PRO:60				
راحت	AJ:238	ADV:53	N:2	RES:1		
راحت‌تر	AJ:61	ADV:43				
راست	AJ:559	N:173	ADV:33	V:2		
راهنما	N:67	AJ:19				
رد	N:803	AJ:287	RES:3			
رسول	N:653	RES:15	AJ:1			
روان	AJ:185	N:134	ADV:4			
روبرو	N:489	AJ:406	P:14			
روزه‌دار	N:28	AJ:27				
روشن	AJ:1957	N:12	ADV:10			
ریاضی	N:329	AJ:17				
رئیس	AJ:34	N:15				
زار	AJ:17	N:12				
زرد	AJ:195	N:19				
زننده	AJ:1159	ADV:12	N:3			

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
زوج	N:122	AJ:17	CL:1			
زیبا	AJ:552	N:78	ADV:22			
زیر	P:2393	AJ:1848	N:730			
زین	P:27	N:23				
ژنتیک	AJ:115	N:67				
ژئوپلیتیک	N:28	AJ:19				
ساده	AJ:691	ADV:21	N:6			
سبز	AJ:1056	N:31				
سخت	AJ:563	ADV:203				
سخنگو	N:78	AJ:11				
سرخ	AJ:503	N:30				
سرکوب	N:198	AJ:28				
سفید	AJ:830	N:32				
سلسله	N:329	CL:92				
سلطان‌نشین	AJ:12	N:11				
سوء	AJ:83	N:67	RES:3			
سیاه	AJ:633	N:30				
شایسته	AJ:601	N:13				
شب	N:2339	ADV:14				
شتابان	AJ:26	ADV:26				
شجاع	AJ:119	N:19				
شدید	AJ:1326	V:55	RES:1	ADV:1		
شروور	N:55	AJ:41				
شفاف	AJ:284	ADV:10				
شگفت	AJ:51	N:12	ADV:2			
شل	N:64	AJ:41				
شناور	AJ:131	N:35				
شهر	N:9072	RES:19				
شیرخوار	AJ:24	N:11				
صاحب‌نظر	AJ:62	N:28				
صادره	AJ:245	N:92				
طرفدار	N:230	AJ:63				
طلا	N:419	AJ:101				
طلای	N:118	AJ:18				

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
طی	P:2739	N:1281	MN:1			
عام	AJ:549	N:40	RES:4			
عتیق	N:24	AJ:11				
عتیقه	AJ:34	N:13				
عجب	DET:44	ADV:41	N:6	AJ:3		
عقب	ADV:306	AJ:93	P:14	N:1		
عقلا	ADV:19	N:11	RES:1			
عکس	N:693	AJ:24				
عمد	AJ:32	N:15	RES:1			
عمیقتر	AJ:102	ADV:17				
عیار	N:32	RES:25	AJ:1			
غالب	AJ:137	DET:59	N:49			
غلط	AJ:341	N:13	ADV:8			
غیر	N:1607	RES:10	AJ:1			
غیر رشید	N:11	AJ:10				
غیر مستقیم	AJ:281	ADV:22				
فارغ‌التحصیل	N:212	AJ:29				
فدای	N:79	AJ:11				
فرار	N:613	AJ:10				
فرد	N:1743	AJ:11				
فلان	DET:292	PRO:30				
فلج	AJ:96	N:23				
فوری	AJ:217	ADV:22				
فی	RES:326	N:14				
فی‌مابین	AJ:32	P:14				
فیلمساز	N:93	AJ:19				
فیما بین	AJ:33	P:23				
قاچاق	N:611	AJ:21				
قال	RES:50	N:11				
قرض‌الحسنه	AJ:56	N:10	ADV:1			
قرمز	AJ:393	N:20				
قضاییه	AJ:296	N:85				
قطره	N:77	CL:22				
قهرمان	N:438	AJ:103				

هم‌تکاه‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
کارشناس	N:495	AJ:12				
کافی	AJ:1395	N:20				
کان	RES:70	N:23	CONJ:9	ADV:1		
کبود	AJ:46	N:16				
کذب	N:44	AJ:17	ADV:1			
کروات	AJ:11	N:10				
کشیدن	N:292	V:10				
کنارش	N:27	P:21				
کنت	N:32	RES:10				
کوچک	AJ:1567	N:23	ADV:5			
کور	AJ:158	N:20	ADV:1			
کوشا	N:35	AJ:13				
گاه	ADV:546	N:535				
گود	AJ:17	N:14				
گونه	CL:3630	N:1096				
لازمه	AJ:123	N:51				
لای	P:50	N:25				
لب	N:237	P:17				
ما	PRO:19684	RES:182				
مادر	N:1225	AJ:43				
مادون	AJ:19	N:13	P:2			
ماده	N:2525	AJ:30				
مارپیچ	N:16	AJ:10				
مازاد	AJ:161	N:64				
ماست	V:533	N:62				
ما فوق	AJ:33	N:19	ADV:2			
ما قبل	P:36	AJ:17	N:2			
مانند	ADV:3161	N:10	V:2			
ماورای	N:47	P:12				
ماهه	N:234	RES:88	V:4			
مبارک	AJ:565	N:96				
مبارکه	N:171	AJ:37				
مینا	N:130	AJ:13				
مثبت	AJ:1077	ADV:40				

هم‌نگاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
مجاز	AJ:466	N:13	ADV:1			
مجدد	AJ:671	ADV:27	N:1			
مجرم	AJ:59	N:46				
مجری	N:257	AJ:11				
محال	AJ:109	N:11				
محض	N:233	AJ:172	P:3			
محکم	AJ:313	ADV:79				
مدار	N:309	V:13				
مدارا	AJ:57	N:35				
مدرسه‌ساز	AJ:57	N:38				
مدعی	AJ:237	N:210				
مرجع	N:378	AJ:15				
مرز‌نما	N:19	AJ:14				
مرمر	AJ:31	N:18				
مريض	AJ:92	N:10				
مزدور	AJ:70	N:14				
مستقیم	AJ:1239	ADV:31	N:1			
مشکل	N:2316	AJ:174	ADV:16			
مشکلی	N:344	AJ:21				
معتاد	AJ:101	N:52				
مفید	AJ:791	N:57	ADV:7	RES:3		
مقدار	N:1286	CL:115				
مگر	ADV:607	CONJ:198	N:2			
ملی‌پوش	AJ:72	N:40				
ممکن	AJ:3523	N:13	ADV:4			
منافی	AJ:24	N:12				
منم	V:29	PRO:20	N:7			
موجود	AJ:3232	N:63	RES:1			
موشک‌انداز	AJ:13	N:11				
مهم	AJ:3339	N:227	ADV:23			
میان	N:3752	P:3453	V:13	AJ:2		
میزان	N:4110	CL:121				
میزبان	N:277	AJ:25				
میمون	AJ:31	N:24				

هم‌تکاره‌ها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
نایغه	N:14	AJ:12				
ناچار	AJ:214	ADV:11				
ناخودآگاه	AJ:47	ADV:27	N:6			
نادر	N:324	AJ:117				
ناشناس	AJ:103	N:24	ADV:1			
ناگزیر	AJ:105	ADV:104				
نامدار	AJ:88	N:12				
نزدیک	AJ:2160	N:312	P:125			
نسخه	N:349	CL:29				
نسیه	N:27	AJ:17	ADV:1			
نظریه‌پرداز	N:12	AJ:11				
نفتکش	N:57	AJ:24				
نقد	N:824	AJ:127	ADV:2			
نقره	N:207	AJ:54				
نگهبان	N:1078	AJ:26				
نمونه	N:931	AJ:266	CL:8	V:1		
نو	AJ:927	N:15	ADV:1			
نوبنیاد	AJ:12	N:10				
نوجوان	N:488	AJ:53				
نوع	N:2119	CL:1609				
نیک	AJ:109	ADV:30	N:24			
نیکو	AJ:88	N:19	ADV:5			
وابسته	AJ:1054	N:38				
واجب	AJ:207	N:17	RES:1			
وارد	P:2410	N:1803	AJ:16			
واگذار	AJ:637	V:20	N:8			
والا	CONJ:161	AJ:80	RES:5	N:5		
والله	RES:25	N:11	ADV:1			
وان	N:50	CONJ:13	RES:3			
ویا	N:28	CONJ:16	CONDET:1			
ور	N:25	CONJ:20	P:2			
ورای	P:41	N:31				
وسط	N:249	P:94	AJ:25	RES:1		
وضوح	AJ:57	N:39				

هم‌نگارها	برچسب ۱	برچسب ۲	برچسب ۳	برچسب ۴	برچسب ۵	برچسب ۶
ولو	CONJ:131	AJ:10	RES:9			
هان	INT:34	RES:11	N:6			
هرچند	CONJ:815	ADV:28				
هم	CONJ:19199	PRO:2001	RES:12	N:9		
همراه	N:2677	AJ:970				
همسایه	N:234	AJ:53				
هوادر	AJ:56	N:16				
هی	RES:78	ADV:13	N:4			
یا	CONJ:20583	RES:23	INT:1			
یتیم	N:45	AJ:26				
یخبندان	N:28	AJ:13				

الهام علایی ابوذر

دانش آموخته دکتری تخصصی در رشته زبان‌شناسی همگانی از دانشگاه تهران است. ایشان هم‌اکنون استادیار پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) است. زبان‌شناسی رایانه‌ای (سیستم‌های تبدیل متن به گفتار، نظام‌های نوشتاری، یادگیری ماشینی و پردازش زبانی)، زبان‌شناسی نظری و پیکره‌ای از جمله علایق پژوهشی وی است.

