

نرم افزارهای نمونه در حوزه تحلیل متن: مزایا و کاربردها برای کاربران نهایی

محمدرضا فلاحتی قدیمی^۱، محمد خسروانیان^۲

^۱ دانشیار گروه پژوهشی زبانشناسی رایانشی مؤسسه استنادی و پایش علم و فناوری جهان اسلام (ISC)، شیراز، ایران.

^۲ دانشجوی کارشناسی ارشد رشته آموزش زبان انگلیسی، دانشگاه خلیج فارس، بوشهر، ایران.

نام نویسنده مسئول:

محمدرضا فلاحتی قدیمی

تاریخ دریافت: ۱۴۰۵/۰۴/۱۷

تاریخ پذیرش: ۱۴۰۵/۰۶/۳۰

چکیده

در سال‌های اخیر، گسترش روزافزون تولید متون دیجیتال، استفاده از ابزارهای برخط تحلیل متن را به بخشی جدایی‌ناپذیر از فرایندهای پژوهشی، آموزشی و تولید محتوا تبدیل کرده است. پیشینه این حوزه از بنیان‌های زبان‌شناسی پیکره‌ای و پردازش آماری زبان آغاز شده و با پژوهش‌هایی در زمینه نمره‌دهی خودکار نوشتار، تحلیل خطای دستوری و بررسی انسجام متن تداوم یافته است. با وجود این سابقه پژوهشی، شکاف قابل توجهی میان یافته‌های نظری و آگاهی کاربران عادی از ابزارهای در دسترس برای تحلیل عملی متن وجود دارد. مقاله حاضر با رویکردی توصیفی-مقایسه‌ای، شش نرم‌افزار برخط و رایگان تحلیل متن، شامل آنالاین یوتیلیتی، تکست آنالایزر، یوزینگ اینگلیش، رود تو گرامر، تکست فیکسر و تکست آنالایسیس آنالاین را بررسی کرده است. برای این منظور، شیوه کاربری، شیوه‌های ورود داده (تایپ مستقیم، بارگذاری فایل یا درونداد نشانی وب‌گاه)، دامنه شاخص‌های آماری و زبانی قابل استخراج و میزان پشتیبانی از زبان فارسی در هر نرم‌افزار بررسی و در قالب جدولی مقایسه‌ای با معیارهای نقاط قوت، نقاط ضعف، هزینه دسترسی و مخاطب پیشنهادی تحلیل شد. یافته‌ها نشان می‌دهد آنالاین یوتیلیتی برای کاربران فارسی‌زبان غیرمتخصص، تکست آنالایزر برای بهینه‌سازی محتوای وب، یوزینگ اینگلیش و رود تو گرامر برای پژوهشگران و مدرسان زبان انگلیسی، تکست فیکسر برای پردازش حجیم اسناد و تکست آنالایسیس آنالاین برای کاربردهای تخصصی پردازش زبان طبیعی مناسب‌ترند. مهم‌ترین محدودیت مشترک این ابزارها، پشتیبانی ناقص یا محدود از زبان فارسی است که عمدتاً به شمارش واژه و کاراکتر منحصر می‌ماند. نتایج این پژوهش می‌تواند راهنمای عملی کاربران در انتخاب ابزار متناسب با نیاز، زبان و سطح تحلیل مورد نظر باشد و ضرورت توسعه منابع پردازش زبان فارسی را برجسته سازد.

واژگان کلیدی: ابزارهای تحلیل متن، تحلیل متن رایانشی، خوانایی و چگالی واژگانی، ارتقاء کیفیت نگارش، نرم‌افزارهای پشتیبان زبان فارسی.

مقدمه

در حوزه آموزش زبان و به طور کلی تألیف و نگارش متن، استفاده از ابزار خودکار برای تصحیح و ارتقاء دستوری، واژگانی و سبکی متن یا نوشته می‌تواند از نظر زمان و هزینه صرفه‌جویی بسیاری ایجاد نماید. ضمناً استفاده از ابزار خودکار تصحیح متن می‌تواند نوعی یکدستی در فرایند تصحیح متن ایجاد کرده و مشکل تفاوت بین ارزیاب‌های انسانی را تا حدودی مرتفع نماید. البته سازندگان نرم‌افزارها و ابزارهای تصحیح متن در تولید محصول خود از رویکردهای مختلفی استفاده می‌کنند. برای نمونه، سینکلر^۱ (۱۹۹۱) با تأکید بر مطالعه زبان بر اساس شواهد واقعی و در مقیاس بزرگ، فلسفه زبان‌شناسی پیکره‌ای را صورتبندی کرد که در آن تمرکز اصلی بر روی شمارش کلمات، محاسبه فراوانی واژگان و استخراج الگوهای همبستگی قرار داشت. یا مانینگ^۲ و شوتره^۳ (۱۹۹۹) در پژوهش خود، الگوهای زبانی را مستقیماً از داده‌های متنی استخراج کردند. به همین ترتیب، دوولین^۴ و همکاران (۲۰۱۹) با معرفی نرم‌افزار پرت^۵ تحلیل متن را از استخراج الگوهای آماری صرف به سمت درک معنایی و استنتاجی سوق دادند.

در طی سال‌های گذشته در حوزه تحلیل خودکار متن، شاهد پیشرفت‌های چشمگیری بوده‌ایم که از آن جمله می‌توان به ابزار نمره‌دهی خودکار نوشتار که توسط کی^۶ و انگ^۷ (۲۰۱۹) ارائه شد، اشاره کرد. در راستای همین کاربردهای آموزشی، پژوهش‌های دیگری نیز به تحلیل جنبه‌های مختلف نوشتار زبان‌آموزان پرداخته‌اند. برای نمونه، جندا^۸ و همکاران (۲۰۱۹) برای نمره‌دهی خودکار انشا مدلی ارائه دادند که به طور همزمان ویژگی‌های نحوی، معنایی و احساسی را در هم تلفیق کرده و از این طریق به دقت بالایی دست می‌یابد. همچنین، ولندری^۹ و هاریدا^{۱۰} (۲۰۲۱) با رویکرد تحلیل خطا، نشان دادند که رایج‌ترین خطاهای دستوری در انشاهای زبان‌آموزان، حرف تعریف، اسم، حرف اضافه و زمان را شامل می‌شود. از سوی دیگر، تحلیل انسجام در نوشتار زبان‌آموزان همواره یکی از موضوعات محوری در تحلیل گفتمان بوده است. سیرگار^{۱۱} و همکاران (۲۰۲۳) در پژوهشی بر روی نوشته‌های دانشجویان انگلیسی به عنوان یک زبان خارجی دریافتند که ارجاع (۵۹ درصد) بیشترین و حذف (۳/۰ درصد) کمترین کاربرد را در میان ابزارهای انسجام دستوری دارد و تکرار (۸۹ درصد) نیز غالب‌ترین ابزار انسجام واژگانی است. همچنین، فلاح‌تی قدیمی فومنی و خسروانیان (۱۴۰۲) مجموعه‌ای از نرم‌افزارهای کاربردی در حوزه تلفظ اسامی خاص را معرفی کرده و به بررسی نقاط قوت و ضعف هر نرم‌افزار پرداختند.

با توجه به پژوهش‌هایی نظیر آنچه در سطور بالا ذکر شد، مقاله حاضر ضمن بیان اهمیت روزافزون این ابزار خودکار درصدد است تا با رویکردی توصیفی و مقایسه‌ای، تصویری شفاف از قابلیت‌ها و محدودیت‌های منتخبی از نرم‌افزارهای حوزه تحلیل متن را ارائه دهد. بر این اساس در این مقاله، نرم‌افزارهای آنلاین یوتیلیتی^{۱۲}، تکست آنالایزر^{۱۳}، یوزینگ اینگلیش^{۱۴}، رود تو گرامر^{۱۵}، تکست فیکسر^{۱۶} و تکست آنالایسیس آنلاین^{۱۷} مورد بررسی قرار خواهد گرفت.

علاوه بر پیشینه ذکرشده، پژوهش‌های اخیر در حوزه ارزیابی خودکار نوشتار نیز شواهد تکمیلی در خصوص کارکرد این دسته از ابزارها ارائه داده‌اند. برای نمونه، دینگ^{۱۸} و زو^{۱۹} (۲۰۲۴) در مروری نظام‌مند بر ۳۹ مقاله مرتبط با سه نرم‌افزار گرامرلی^{۲۰}،

1. Sinclair

2. Manning

3. Schütze

4. Devlin

5. Bidirectional Encoder Representations from Transformers (BERT)

6. Ke

7. Ng

8. Janda

9. Wulandari

10. Harida

11. Siregar

12. Online Utility

13. Textalyser

14. Usingenglish

15. Road To Grammar

16. Text fixer

17. Text Analysis Online

18. Ding

پیگای^{۲۱} و کرایتریون^{۲۲} نشان دادند که این ابزارها اثر مثبتی بر ارتقاء توانایی نگارش زبان‌آموزان داشته و هم زبان‌آموزان و هم مدرسان نگرش مساعدی نسبت به آنها دارند؛ هرچند نویسندگان بر ضرورت بررسی بیشتر عملکرد این سامانه‌ها در عصر هوش مصنوعی مولد تأکید کردند. در پژوهشی دیگر، دیزون^{۲۳} و گیید^{۲۴} (۲۰۲۱) با استفاده از طرح متعادل‌شده و بررسی ۳۱ زبان‌آموز ژاپنی فراگیر زبان انگلیسی به عنوان زبان خارجی دریافتند که استفاده از نرم‌افزار گرامرلی به‌طور معناداری دقت دستوری و غنای واژگانی نوشتار را بهبود می‌بخشد، در حالی که تأثیر معناداری بر روانی نوشتار و پیچیدگی نحوی مشاهده نشد. همچنین نونس^{۲۵} و همکاران (۲۰۲۱) در مروری نظام‌مند بر پژوهش‌های سال‌های ۲۰۰۰ تا ۲۰۲۰ در محیط‌های مدرسه‌ای، اثربخشی شش نرم‌افزار ارزیابی خودکار نوشتار را بر روی ۱۶۵۹ دانش‌آموز بررسی کردند و دریافتند که در اکثر موارد بازخورد خودکار دست‌کم بر یکی از شاخص‌های مرتبط با نوشتار تأثیر مثبتی داشته است.

از منظر آموزشی نیز، الحربی^{۲۶} (۲۰۲۳) در مروری بر ۱۰۴ مقاله علمی-پژوهشی نشان داد که ابزارهای نگارشی مبتنی بر هوش مصنوعی به‌طور فزاینده‌ای در کلاس‌های زبان خارجی به کار گرفته می‌شوند، اما یافته‌های موجود درباره میزان اثربخشی آنها گاه ناهمسو است و نیازمند بررسی دقیق‌تری هستند. فو^{۲۷} و همکاران (۲۰۲۲) نیز در تحلیل ۴۸ مقاله پیرامون بازخورد ابزارهای ارزیابی خودکار نوشتار، انواع بازخورد، بازده‌های یادگیری و پیامدهای آموزشی این فناوری را دسته‌بندی کرده و بر غلبه رویکردهای کمی در این حوزه پژوهشی تأکید کردند. در سطحی گسترده‌تر، دیوا^{۲۸} و همکاران (۲۰۲۰) با ارائه چارچوبی طبقه‌بندی‌شده از سامانه‌های بازخورد خودکار یادگیرندگان، چالش‌ها و فرصت‌های پیش روی این فناوری‌ها را بررسی کرده و بر ضرورت طراحی سامانه‌هایی با قابلیت تعمیم به حوزه‌های محتوایی و زبانی متنوع تأکید نمودند. در زمینه سنجش کیفیت متن نیز، السدیس^{۲۹} (۲۰۲۵) با تحلیل روند خوانایی چکیده‌های علمی طی سال‌های اخیر دریافت که پس از انتشار ابزارهای هوش مصنوعی مولد تغییری محسوس در سطح خوانایی متون علمی رخ داده است که می‌تواند بر کاربرد شاخص‌های خوانایی در نرم‌افزارهای تحلیل متن نیز اثرگذار باشد.

علاوه بر رویکردهای ارزیابی نوشتار، بخشی از پژوهش‌های اخیر بر شاخص‌های کمی زبانی نظیر چگالی و پیچیدگی واژگانی متمرکز بوده‌اند. برای نمونه، ژو^{۳۰}، وانگ^{۳۱} و پانگ^{۳۲} (۲۰۲۴) با تحلیل پیکره‌ای بزرگ‌مقیاس شامل چکیده‌های مقالات پژوهشی از ۱۶ رشته علمی طی سال‌های گذشته دریافتند که چگالی واژگانی متون علمی روندی افزایشی داشته و رشته‌های علوم سخت نسبت به علوم نرم چگالی بالاتری نشان می‌دهند. در همین راستا، لو^{۳۳} (۲۰۱۲) با طراحی سامانه‌ای محاسباتی برای اندازه‌گیری خودکار سه بعد غنای واژگانی (یعنی چگالی، پیچیدگی و تنوع واژگانی) نشان داد که این شاخص‌ها با قضاوت ارزیابان درباره کیفیت روایت‌های شفاهی زبان‌آموزان زبان دوم همبستگی معناداری دارند.

از منظر تحلیل انسجام متن نیز، گریسر^{۳۴} و همکاران (۲۰۰۴) با معرفی نرم‌افزار کوه‌متری‌کس، ابزاری رایانه‌ای ارائه کردند که بیش از ۲۰۰ شاخص مرتبط با انسجام، زبان و خوانایی متن را اندازه‌گیری می‌کند و بر خلاف فرمول‌های سنتی خوانایی، به روابط انسجامی، دانش جهانی و ویژگی‌های گفتمانی متن نیز حساس است. کراسلی^{۳۵}، کایل^{۳۶} و مک‌نامارا^{۳۷} (۲۰۱۶) نیز با

19. Zou

20. Grammarly

21. Pigai

22. Criterion

23. Dizon

24. Gayed

25. Nunes

26. Alharbi

27. Fu

28. Deeva

29. Alsudais

30. Zhu

31. Wang

32. Pang

33. Lu

34. Graesser

35. Crossley

36. Kyle

توسعه ابزار خودکار تحلیل انسجام متن، رشد به‌کارگیری ادوات انسجامی در نوشتار زبان‌آموزان زبان دوم را طی یک نیم‌سال تحصیلی بررسی و رابطه معناداری میان این شاخص‌ها و قضاوت کیفیت مقاله گزارش کردند. در پژوهشی دیگر، ژو^{۳۸} و همکاران (۲۰۲۳) با استفاده از کوه‌متریکس عملکرد نوشتاری چت جی پی تی^{۳۹} را با زبان‌آموزان چینی سطح متوسط زبان انگلیسی مقایسه کردند و دریافتند اگرچه چت جی پی تی در روایت‌پذیری و انسجام ارجاعی عملکرد بهتری داشت، در انسجام عمیق و سادگی نحوی همچنان از زبان‌آموزان انسانی عقب بود. یافته‌ای که پچپرسرت^{۴۰} (۲۰۲۱) نیز با استفاده از همین ابزار در ارزیابی نوشتار زبان‌آموزان زبان خارجی تأیید کرده و نشان داد طول متن و میزان انسجام ارجاعی و عمیق از عوامل مؤثر بر عملکرد نوشتاری است. در کنار این ابزارهای تخصصی، استیس^{۴۱} و همکاران (۲۰۲۴) در پژوهشی مقایسه‌ای دریافتند که ارزیابان انسانی آموزش‌دیده همچنان کیفیت بازخورد بالاتری نسبت به چت جی پی تی ارائه می‌دهند، هرچند این ابزار در برخی زمینه‌ها می‌تواند مکمل مفیدی برای بازخورد نوشتاری باشد. سرانجام، تاکاکی^{۴۲} و دوترا^{۴۳} (۲۰۲۴) در مروری نظام‌مند بر ۵۲ پژوهش نشان دادند که فنون داده‌کاوی متن و پردازش زبان طبیعی در آموزش عالی از راه دور عمدتاً به نفع مدرسان طراحی شده‌اند و جای خالی پژوهش‌های متمرکز بر دانشجویان و مدیران آموزشی همچنان محسوس است؛ نکته‌ای که بر ضرورت طراحی کاربرمحور ابزارهای تحلیل متن، از جمله در حوزه زبان فارسی تأکید می‌کند.

نرم‌افزارهای نمونه تحلیل متن

نرم‌افزار آنلاین بوتیلیتی

این نرم‌افزار که از طریق آدرس <https://www.online-utility.org/text/analyzer.jsp> قابل دسترسی است، یکی از کاربرپسندترین ابزارهای تحلیل‌گر متن محسوب می‌شود و از ویژگی‌های برجسته آن پشتیبانی کامل از زبان فارسی است که آن را به گزینه‌ای مناسب برای کاربران ایرانی تبدیل کرده است. کاربر پس از ورود به صفحه اصلی، می‌تواند متن مورد نظر خود را در جعبه مربوطه وارد کند یا با استفاده از گزینه بارگذاری فایل، یک فایل متنی را مستقیماً به نرم‌افزار معرفی نماید؛ سپس با کلیک زدن بر روی دکمه Process text، فرایند پردازش آغاز می‌شود. خروجی این نرم‌افزار در قالب یک جدول خلاصه ارائه می‌گردد که اطلاعات آماری متنوعی از جمله تعداد کاراکترها (همراه با فاصله یا بدون احتساب آن)، تعداد کلمات، چگالی واژگانی (که نسبت واژه‌های متمایز به کل رخداد واژه‌ها را نشان می‌دهد)، تعداد جملات و تعداد سیلاب‌های متن را شامل می‌شود. افزون بر این آمارهای پایه، نرم‌افزار قابلیت استخراج فراوان‌ترین عبارات چندکلمه‌ای موجود در متن را نیز دارد و در بخش دیگری از خروجی، فهرست کاملی از تک‌واژه‌های متن به همراه فراوانی رخداد هر یک از آنها، به‌صورت مرتب‌شده از پربسامدترین به کم‌بسامدترین واژه، در اختیار کاربر قرار می‌دهد تا بتواند تحلیل دقیق‌تری از توزیع واژگانی متن خود داشته باشد.

37. McNamara

38. Zhou

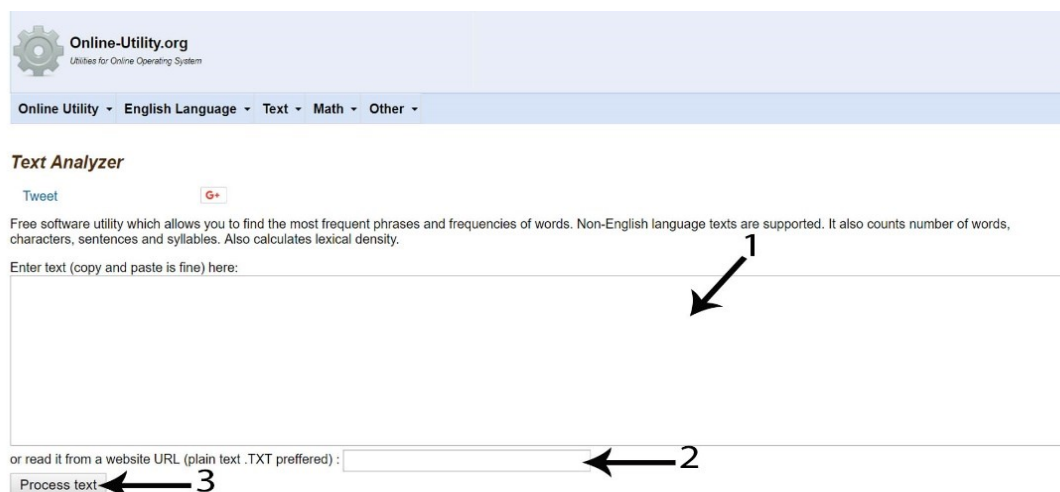
39. ChatGPT

40. Petchprasert

41. Steiss

42. Takaki

43. Dutra



Online-Utility.org
Utilities for Online Operating System

Online Utility ▾ English Language ▾ Text ▾ Math ▾ Other ▾

Text Analyzer

[Tweet](#) [G+](#)

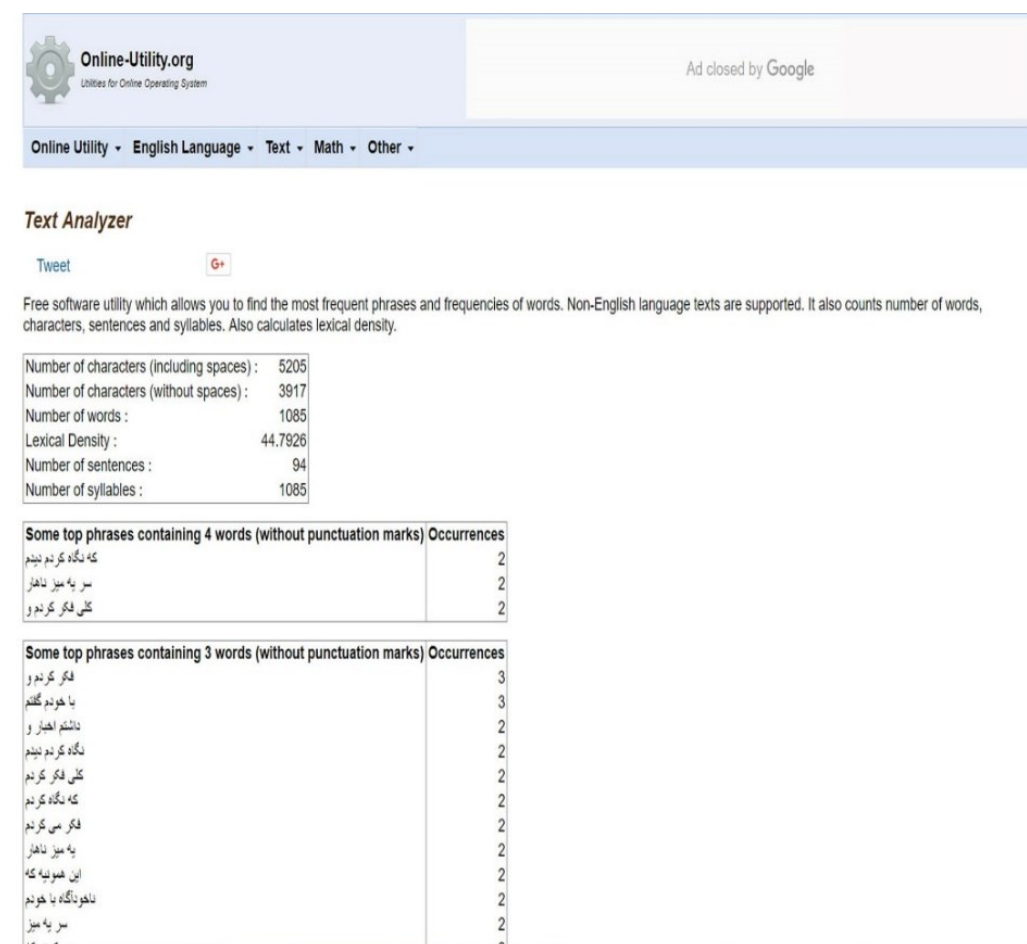
Free software utility which allows you to find the most frequent phrases and frequencies of words. Non-English language texts are supported. It also counts number of words, characters, sentences and syllables. Also calculates lexical density.

Enter text (copy and paste is fine) here:

or read it from a website URL (plain text .TXT preferred):

Process text

شکل ۱: صفحه اصلی نرم افزار آنلاین یوتیلیتی.



Online-Utility.org
Utilities for Online Operating System

Online Utility ▾ English Language ▾ Text ▾ Math ▾ Other ▾

Text Analyzer

[Tweet](#) [G+](#)

Free software utility which allows you to find the most frequent phrases and frequencies of words. Non-English language texts are supported. It also counts number of words, characters, sentences and syllables. Also calculates lexical density.

Number of characters (including spaces) :	5205
Number of characters (without spaces) :	3917
Number of words :	1085
Lexical Density :	44.7926
Number of sentences :	94
Number of syllables :	1085

Some top phrases containing 4 words (without punctuation marks)	Occurrences
که نگاه کردم دینیم	2
سر په میز ناهار	2
کشی فکر کردم و	2

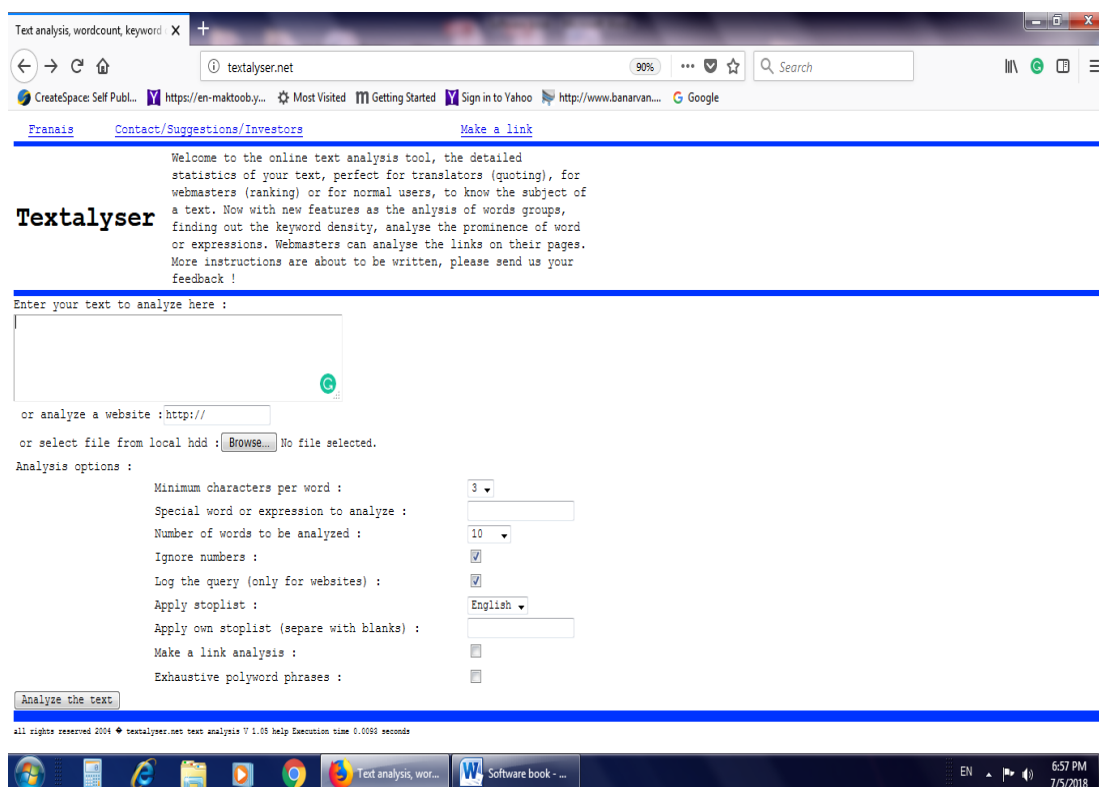
Some top phrases containing 3 words (without punctuation marks)	Occurrences
فکر کردم و	3
با خوردم گفتم	3
داشتم اختیار و	2
نگاه کردم دینیم	2
کشی فکر کردم	2
که نگاه کردم	2
فکر می کردم	2
په میز ناهار	2
این هوانیه که	2
ناخودگاه با خوردم	2
سر په میز	2

شکل ۲: نحوه عملکرد نرم افزار آنلاین یوتیلیتی.

نرم افزار تکست آنالایزر

نرم افزار تکست آنالایزر با آدرس اینترنتی <http://textalyser.net/> یکی دیگر از ابزارهای برخط تحلیل متن است که با ارائه پردازش‌های آماری متعدد، نیازهای کاربران را در زمینه بررسی کمی متون انگلیسی برآورده می‌سازد. لازم به ذکر است که این نرم افزار متأسفانه از زبان فارسی پشتیبانی نمی‌کند و بنابراین دامنه کاربرد آن به متون انگلیسی محدود می‌شود. کاربر برای استفاده از این ابزار، پس از ورود به صفحه اصلی، سه روش برای وارد کردن داده در اختیار دارد: (۱) می‌تواند متن مورد نظر را

مستقیماً در جعبه‌ی مربوطه تایپ کند یا آن را از رایانه شخصی برداشته و در جعبه متن قرار دهد. ۲) کاربر همچنین می‌تواند آدرس یک وب‌گاه را در بخش Website وارد نماید تا نرم‌افزار به‌طور خودکار محتوای متنی آن صفحه را استخراج و تحلیل کند، یا ۳) با استفاده از گزینه Browse، یک فایل متنی از پیش آماده را از رایانه خود بارگذاری نماید. پس از انتخاب روش ورود داده، کاربر می‌تواند از طریق گزینه‌های Analysis Options، نوع پردازش‌های مورد نیاز خود را انتخاب کرده و سپس با کلیک‌زدن بر روی دکمه Analyze the Text، فرایند تحلیل را آغاز کند. خروجی نرم‌افزار طیف وسیعی از اطلاعات آماری را شامل می‌شود که از آن جمله می‌توان به تعداد واژه‌های متمایز، تعداد کل واژه‌ها، چگالی واژگانی (نسبت واژه‌های متمایز به کل رخدادهای واژگانی)، درجه خوانایی متن (که نشان‌دهنده سطح دشواری درک مطلب برای مخاطب است)، تعداد جملات، عبارات چندکلمه‌ای، ده واژه برتر از نظر فراوانی و همچنین تعداد کاراکترها و سیلاب‌های متن به همراه درصد هر یک اشاره کرد که همگی در قالبی منظم و قابل‌فهم به کاربر ارائه می‌شوند.



شکل ۳: صفحه اصلی نرم افزار تکست آنالایزر.

**Textalyser
Results**

The complete results, including complexity factor, and other features

Total word count :	234
Number of different words :	176
Complexity factor (Lexical Density) :	75.2%
Readability (Gunning-Fog Index) : (6-easy 20-hard)	12.9
Total number of characters :	2294
Number of characters without spaces :	1467
Average Syllables per Word :	1.69
Sentence count :	14
Average sentence length (words) :	27.07
Max sentence length (words) :	47
([9] in addition schexnayder stating that in truth though stories is a release that is easy to listen to and just a little bit too easy to forget and this is why the album was for those who are seeking a fun but simple encouraging pop rock album)	
Min sentence length (words) :	6
([6] at christian broadcasting network jennifer e)	
Readability (Alternative) beta : (100-easy 20-hard, optimal 60-70)	36.7

Frequency and top words :

شکل ۴: نحوه عملکرد نرم افزار تکست آنالایزر.

نرم افزار یوزینگ اینگلیش

نرم افزار یوزینگ اینگلیش که از طریق آدرس <https://www.usingenglish.com/resources/text-statistics.php> در دسترس کاربران قرار گرفته است، ابزاری کارآمد برای تحلیل آماری متون به شمار می‌رود و قابلیت‌های متنوعی را در زمینه استخراج شاخص‌های کمی و کیفی متن ارائه می‌دهد. این نرم‌افزار می‌تواند پردازش‌های آماری متعددی را انجام دهد که از جمله آنها می‌توان به تعداد واژه‌های متمایز، تعداد کل واژه‌های متن، تعداد جملات، میانگین طول جملات (که نشان‌دهنده پیچیدگی ساختاری متن است)، تعداد بندها، چگالی واژگانی، درصد کلمات دشوار موجود در متن و شاخص تعیین سختی متن اشاره کرد. کاربر پس از ورود به صفحه اصلی نرم‌افزار، باید متن خود را در جعبه تعیین‌شده وارد کرده و سپس با کلیک‌زدن بر روی دکمه Calculate Now!، منتظر بماند تا فرایند پردازش تکمیل شود. خروجی نرم‌افزار در سه بخش مجزا به کاربر نمایش داده می‌شود: بخش نخست با عنوان Text Statistics شامل کلیه آمارهای توصیفی از جمله تعداد واژه‌ها، جملات، بندها و شاخص‌های پیچیدگی است؛ بخش دوم با عنوان Word Length Breakdown به تحلیل و تفکیک کلمات بر اساس طول آنها می‌پردازد و توزیع طول کلمات را به صورت درصدی نشان می‌دهد؛ و بخش سوم با عنوان Word Frequency Cloud، واژه‌های موجود در متن را به شکل ابری نمایش می‌دهد که در آن واژه‌ها به صورت الفبایی مرتب شده‌اند و واژه‌های پربسامدتر با اندازه بزرگ‌تری نسبت به واژه‌های کم‌بسامدتر تایپ شده‌اند تا کاربر بتواند به سرعت مهم‌ترین و پرتکرارترین واژه‌های متن خود را شناسایی کند.

Home > Resources > Text Analyser

Text Content Analysis Tool

Text Analyser Word Analyser

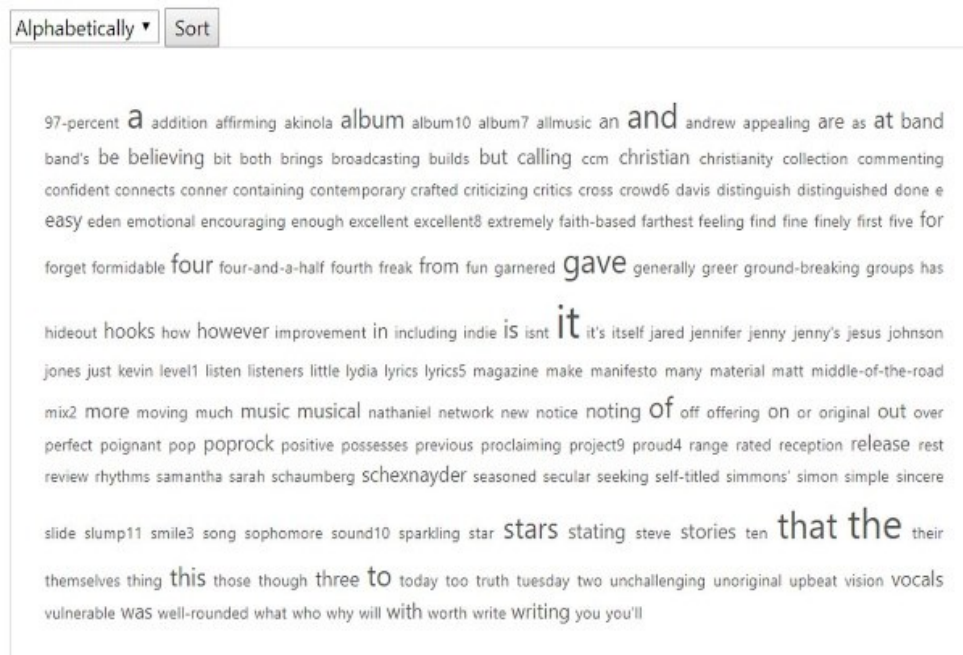
Analyse text content using our free text analysis tool which gives you statistics about a text including word count; unique words; number of sentences; average words per sentence; lexical density; and the Gunning Fog readability index. More detailed statistics are available to members.

Please write or paste the text you would like to analyse in the box below. This analyser will accept text up to 10,000 characters (members can analyse longer texts using our advanced text analyser):

شکل ۵: صفحه اصلی نرم افزار یوزینگ اینگلیش.

Word Frequency Cloud

The **word frequency cloud** shows each word in a bigger or smaller text size, depending on how many times it is found in the text. The bigger the word, the more frequently it occurs. Very common words have been removed to allow you to see the key words in the text. Note: The results work better with longer texts.

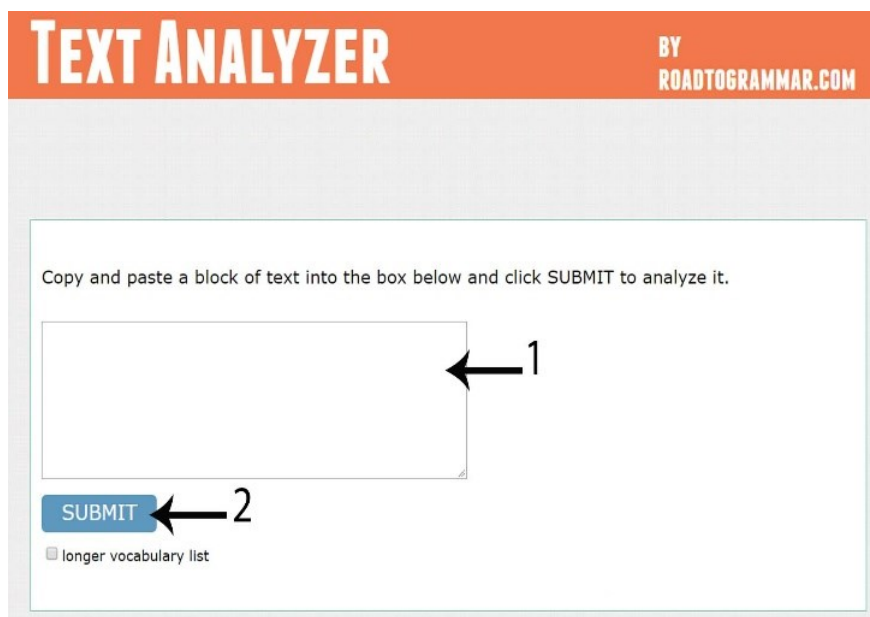


شکل ۶: نمایش عملکرد نرم افزار یوزینگ اینگلیش.

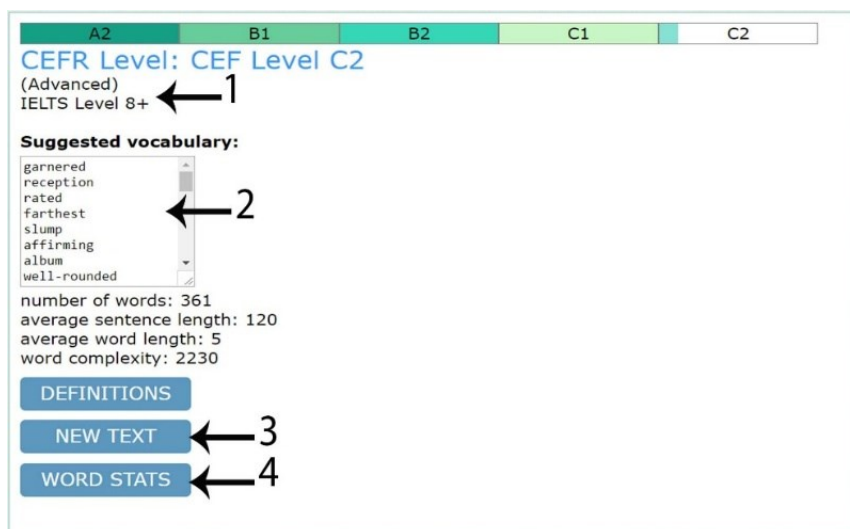
نرم افزار رود تو گرامر

نرم افزار رود تو گرامر با آدرس اینترنتی <http://www.roadtogrammar.com/textanalysis/> یکی از ابزارهای تخصصی تحلیل متن با رویکردی عمدتاً آموزشی و زبان آموزی محسوب می شود که امکانات منحصر به فردی را برای زبان آموزان و مدرسان زبان انگلیسی فراهم می آورد. این نرم افزار قادر است محاسبات آماری پایه ای مانند تعداد کلمات، میانگین طول جملات، میانگین طول کلمات و همچنین شاخص دشواری واژگانی متن را به دقت محاسبه و در اختیار کاربر قرار دهد. اما از ویژگی های متمایز و برجسته این نرم افزار می توان به دو قابلیت ویژه اشاره کرد: نخست، ارائه فهرستی از کلمات دشوار متن با عنوان Suggested

Vocabulary که می‌تواند در فرایند یادگیری و گسترش دایره واژگان زبان‌آموزان نقشی کلیدی ایفا کند؛ و دوم، ارزیابی و نمره‌دهی متن بر اساس معیارهای استاندارد آزمون آیلتس (IELTS) که این ویژگی آن را به ابزاری ارزشمند برای زبان‌آموزانی که قصد شرکت در این آزمون را دارند تبدیل کرده است. کاربر برای استفاده از این نرم‌افزار، پس از ورود به صفحه اصلی، باید متن مورد نظر خود را در جعبه تعیین‌شده وارد کند و سپس با کلیک‌زدن بر روی دکمه SUBMIT، فرایند تحلیل را آغاز نماید. خروجی نرم‌افزار در صفحه جدیدی نمایش داده می‌شود که در بالای آن، سطح کلی متن بر اساس معیار آیلتس به کاربر نشان داده می‌شود؛ در بخش Suggested Vocabulary، فهرست واژگان دشوار به همراه پیشنهادهایی برای یادگیری ارائه می‌گردد؛ و در بخش WORD STATS، درجه سختی تک تک کلمات متن نمایش داده می‌شود.



شکل ۷: صفحه اصلی نرم افزار رود تو گرامر.

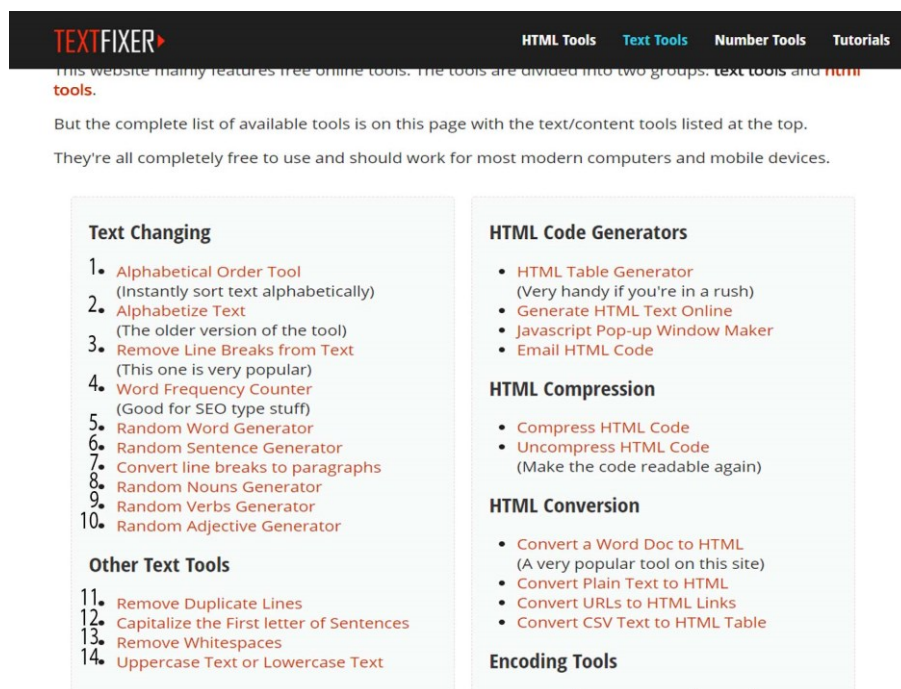


شکل ۸: نحوه عملکرد نرم‌افزار رود تو گرامر.

نرم‌افزار تکست فیکسر

پایگاه اینترنتی تکست فیکسر با آدرس <https://www.textfixer.com>، یک مجموعه جامع و متنوع از ابزارهای برخط پردازش متن را در اختیار کاربران قرار می‌دهد که همگی به‌صورت کاملاً رایگان و بدون نیاز به نصب نرم‌افزار، بر روی رایانه‌های

شخصی و تلفن‌های همراه قابل استفاده هستند. این ابزارها در دو گروه عمده Text Tools (ابزارهای متنی) و HTML Tools (ابزارهای مرتبط با کدهای اچ‌تی‌ام‌ال) دسته‌بندی می‌شوند که گروه نخست شامل ۱۴ ابزار ساده اما کاربردی برای پردازش و تحلیل متون است. کاربر پس از ورود به صفحه اصلی و انتخاب گزینه Text Tools از منوی بالا، با فهرستی از ابزارهای متنوع مواجه می‌شود که هریک کارکرد خاصی دارند: ابزار مرتب‌سازی الفبایی (Alphabetical Order Tool) که می‌تواند فهرستی الفبایی از واژه‌های متن، با امکان حذف کلمات تکراری یا علائم نشانه‌گذاری، تولید کند؛ ابزار شمارشگر فراوانی واژگان (Word Frequency Counter) که یکی از کاربردی‌ترین ابزارهای این مجموعه به‌شمار می‌رود و می‌تواند فهرست کامل واژه‌های متن را به همراه تعداد رخداد هر یک ارائه دهد و از زبان فارسی نیز پشتیبانی می‌کند؛ ابزار حذف خطوط اضافی (Remove Line Breaks) که متن‌های به‌هم‌ریخته را به حالت اولیه بازمی‌گرداند؛ ابزار تولید تصادفی واژه، اسم و فعل که برای مقاصد آموزشی و تقویت دایره واژگان بسیار مفید است؛ و همچنین ابزارهای حذف فضای خالی اضافی، بزرگ‌سازی حروف ابتدای جملات، تغییر حالت بزرگ و کوچک حروف، یکدست‌سازی فاصله بین بندها و تولید جمله تصادفی. کاربر به‌سادگی با وارد کردن متن خود در جعبه مربوطه و انتخاب ابزار مورد نظر، می‌تواند خروجی پردازش‌شده را در کمترین زمان ممکن دریافت نماید؛ این سادگی استفاده، همراه با تنوع ابزارها، تکست فیکسر را به یکی از گزینه‌های مناسب برای کاربرانی تبدیل کرده است که نیاز به پردازش‌های سریع و مقدماتی بر روی متون خود دارند.



شکل ۹: صفحه اصلی پایگاه اینترنتی تکست فیکسر.

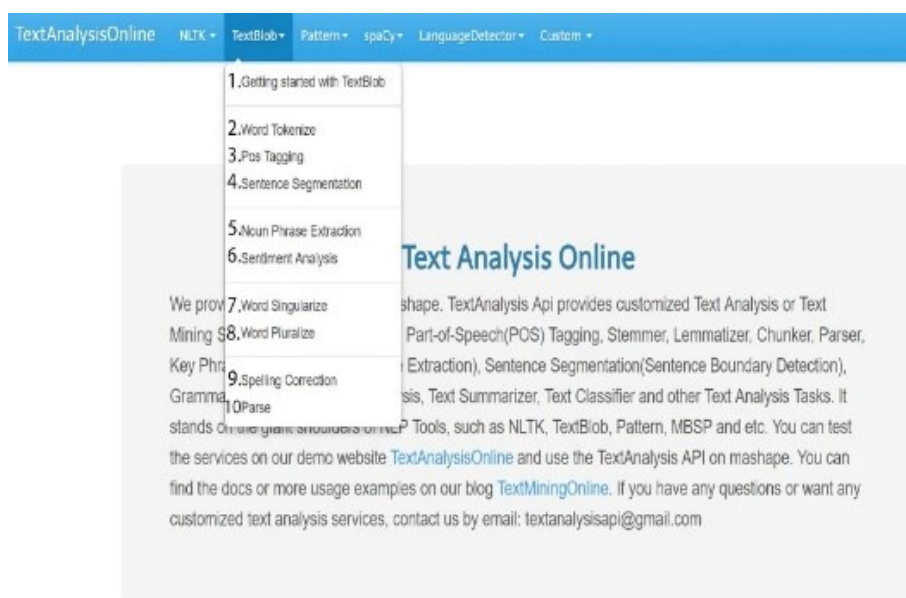
نرم‌افزار تکست آنالیز آنلاین

نرم‌افزار تکست آنالیز آنلاین که از طریق آدرس <http://textanalysisonline.com/> قابل دسترسی است، بی‌شک جامع‌ترین و پیشرفته‌ترین ابزار در میان نرم‌افزارهای معرفی‌شده این مقاله به‌شمار می‌رود؛ چراکه مجموعه‌ای از کتابخانه‌های معروف و قدرتمند پردازش زبان طبیعی در پایتون را در قالب یک رابط کاربری برخط و آسان، گرد هم آورده است. این نرم‌افزار در شش بخش مجزا، خدمات متنوعی را ارائه می‌دهد: بخش نخست بر پایه کتابخانه NLTK (Natural Language Toolkit) است که امکاناتی چون تقطیع واژگانی (تجزیه متن به کلمات مجزا)، تقطیع جمله‌ای (تقسیم متن به جملات)، برچسب‌زنی نحوی واژگان (POS Tagging)، ریشه‌یابی با سه الگوریتم معروف پورتر^{۴۴}، لَنکستر^{۴۵} و اسنوبال^{۴۶} و همچنین ریشه‌یابی کلمات

^{۴۴}. Porter

^{۴۵}. Lancaster

با استفاده از بانک واژگانی وردنت^{۴۶} را در اختیار کاربر قرار می‌دهد. بخش دوم مبتنی بر کتابخانه TextBlob است که علاوه بر تقطیع و برچسب‌زنی، قابلیت‌های پیشرفته‌تری مانند استخراج گروه‌های اسمی موجود در متن، تبدیل کلمات جمع به مفرد و مفرد به جمع، شناسایی و اصلاح خودکار خطاهای املائی و انجام تحلیل نحوی بر روی جملات را ارائه می‌کند. بخش سوم با استفاده از کتابخانه Pattern، علاوه بر امکانات پایه، خدمات پیشرفته‌تری چون تحلیل احساسات (که برای هر جمله حس مثبت یا منفی بودن آن را مشخص می‌کند)، تشخیص وجه جمله و تبدیل صفات ساده به تفضیلی و عالی را فراهم می‌آورد. بخش چهارم بر پایه کتابخانه spaCy که یکی از قدرتمندترین کتابخانه‌های پردازش زبان طبیعی محسوب می‌شود، قابلیت‌هایی نظیر ریشه‌یابی دقیق، برچسب‌زنی نحوی، تشخیص انواع اسمی خاص و همچنین استخراج کلیه اسم‌های موجود در متن را به کاربر ارائه می‌دهد. بخش پنجم با عنوان LanguageDetector، این امکان را فراهم می‌کند که کاربر با وارد کردن یک متن به هر زبانی، زبان آن را تشخیص دهد. نهایتاً بخش ششم با عنوان Custom، مجموعه‌ای از خدمات سفارشی را شامل می‌شود که عبارتند از: شناسایی کلیدواژه‌ها و اصطلاحات کلیدی متن، خلاصه‌سازی خودکار متون طولانی، تحلیل احساسات با ارائه درصد حس مثبت یا منفی و محاسبه درصد شباهت بین دو متن که کاربر می‌تواند با وارد کردن دو متن مجزا، عددی برای میزان شباهت آنها دریافت کند. این تنوع و جامعیت، این نرم‌افزار را برای پژوهشگران، مدرسان و دانشجویانی که به تحلیل‌های عمیق‌تر و تخصصی‌تر نیاز دارند، به ابزاری مطلوب تبدیل می‌کند.



شکل ۱۰: نمای کلی منوی اصلی نرم‌افزار تکست آنالیز آنلاین.

^{۴۶}. Snowball

^{۴۷}. WordNet



شکل ۱۱: نمونه خروجی بخش Custom در نرم‌افزار تکست آنالیز آنلاین.

خلاصه یافته‌ها

برای تسهیل آشنایی مخاطبان با این نرم‌افزارها، خلاصه ویژگی‌های آنها در قالب جدول زیر آورده شده است.

جدول ۱. جمع‌بندی مقایسه‌ای نرم‌افزارهای تحلیل متن.

نرم‌افزار	پشتیبانی از فارسی	نوع تحلیل غالب	نقاط قوت اصلی	نقاط ضعف اصلی	هزینه / دسترسی	مخاطب پیشنهادی
آنلاین یوتیلتی	بله (شمارش واژه/کاراکتر)	آماري- توصیفی	رابط کاربری ساده، ورودی وب‌گاه	فاقد تحلیل‌های عمیق زبانی	رایگان، بدون نیاز به ثبت‌نام	کاربران عمومی، دانشجویان، پژوهشگرانی که نیاز به تحلیل آماری متون فارسی دارند
تکست آنالیزر	خیر (فقط انگلیسی)	بهینه‌سازی محتوا	تحلیل تراکم کلیدواژه، ضریب خوانش وب، گزارش فوری در مرورگر	فاقد تحلیل‌های زبان‌شناختی و پیکره‌ای	رایگان	تولیدکنندگان محتوا، متخصصان سئو، متخصصان وب
یوزینگ اینگلیش	خیر (فقط انگلیسی)	خوانایی و چگالی واژگانی	گسترده‌ترین شاخص‌های خوانش متن	نیاز به ثبت‌نام برای امکانات پیشرفته، عدم تحلیل معنایی	رایگان، ثبت‌نام برای امکانات پیشرفته الزامی است	پژوهشگران زبان‌شناسی، معلمان زبان، نویسندگان متون دانشگاهی
رود تو گرامر	خیر (فقط انگلیسی)	آموزش زبان	برآورد سطح زبانی A1 تا C2، شناسایی واژگان دشوار	نیاز به حداقل ۵۰ واژه برای نتیجه قابل اعتماد	رایگان	مدرسان زبان انگلیسی، تهیه‌کنندگان محتوای آموزشی
تکست فیکسر	بله (شمارش واژه/کاراکتر)	فراوانی و پردازش فایل	پشتیبانی از PDF، DOCX خروجی CSV، ابزارهای PDF، متنوع پردازش متن	فاقد فرمول‌های خوانایی و تحلیل دستوری	رایگان	کاربرانی که با اسناد حجیم کار می‌کنند و نیاز به گزارش خروجی دارند

تکست آنالیز آنالیز	محدود (برخی بخش‌ها)	پردازش زبان طبیعی	گسترده‌ترین عملیات فنی	سازگاری کم بین بخش‌ها، بیشتر نمایشی تا تحلیلی	نمایش‌های آموزشی رایگان	دانشجویان و پژوهشگرانی که قصد آشنایی با مفاهیم پردازش زبان طبیعی را دارند
--------------------------	------------------------	----------------------	---------------------------	---	----------------------------	--

بحث و نتیجه‌گیری

این پژوهش با بررسی شش نرم‌افزار برخط تحلیل متن در طیفی از ابزارهای ساده تا پلتفرم‌های پیشرفته، نشان می‌دهد که انتخاب ابزار مناسب کاملاً به نیاز کاربر، نوع متن، زبان موردنظر و سطح تحلیل بستگی دارد. یافته‌ها حاکی از آن است که اگرچه تمامی این ابزارها قابلیت‌های پایه‌ای همچون شمارش واژه‌ها، کاراکترها و جملات را ارائه می‌دهند، اما در عمق و کیفیت تحلیل‌های زبانی، پشتیبانی از زبان‌های مختلف و کاربردهای تخصصی تفاوت‌های چشمگیری با یکدیگر دارند. برای تحلیل‌های آماری ساده و سریع، نرم‌افزارهایی مانند آنالیز یوتیلیتی با پشتیبانی از زبان فارسی و رابط کاربری آسان، گزینه‌ای مناسب برای کاربران غیرمتخصص به‌شمار می‌روند. با این حال، این دسته از ابزارها عمدتاً به تحلیل‌های سطحی و توصیفی محدود بوده و فاقد قابلیت‌های پیشرفته‌تری مانند برچسب‌زنی نحوی، تحلیل احساسات یا طبقه‌بندی سطح زبانی‌اند. در مقابل، ابزارهایی مانند یوزینگ اینگلیش و رود تو گرامر که به‌طور خاص برای متون انگلیسی طراحی شده‌اند، دامنه وسیع‌تری از شاخص‌های خوانایی، چگالی واژگانی، فرمول‌های سنجش سختی متن و حتی برآورد سطح زبانی متن را ارائه می‌دهند که این ویژگی‌ها آنها را به ابزارهایی مطلوب برای معلمان زبان، پژوهشگران آموزش زبان و تهیه‌کنندگان محتوای آموزشی تبدیل می‌کند.

از سوی دیگر، نرم‌افزار تکست آنالیز با تمرکز بر بررسی تراکم کلیدواژه‌ها و تکرار عبارات، بیشتر برای بهینه‌سازی محتوای وب‌گاه‌ها کاربرد دارد، هرچند فاقد عمق تحلیل‌های زبان‌شناختی است. نرم‌افزار تکست فیکس نیز با پشتیبانی از بارگذاری فایل‌های متنی، پیدی اف و قابلیت خروجی گرفتن در قالب‌های مختلف، برای استخراج داده‌های فراوانی‌محور از اسناد طولانی مناسب‌تر است. در رأس این طیف، نرم‌افزار تکست آنالیز spaCy و NLTK، TextBlob، Pattern و spaCy وسیع‌ترین مجموعه از عملیات‌های فنی از جمله تقطیع، ریشه‌یابی، برچسب‌زنی نحوی، تحلیل احساسات و اسامی خاص و خلاصه‌سازی خودکار را در اختیار کاربر قرار می‌دهد.

از چالش‌های مهمی که در این بررسی مشاهده شد، مسئله پشتیبانی از زبان‌های غیرانگلیسی، به‌خصوص زبان فارسی است. در حالی که ابزارهایی مانند آنالیز یوتیلیتی و تکست فیکس تا حدی از فارسی پشتیبانی می‌کنند، اما این پشتیبانی عمدتاً به شمارش واژه‌ها و کاراکترها محدود است و فرمول‌های خوانایی، فهرست‌های واژگانی، ریشه‌یابی، یا تحلیل‌های دستوری همچنان برای زبان انگلیسی طراحی شده‌اند. این شکاف زبانی، که در پیشینه پژوهش نیز به آن اشاره شد، نشان‌دهنده ضرورت سرمایه‌گذاری بر روی توسعه منابع و ابزارهای پردازش زبان فارسی است. به‌طور کلی، این مطالعه تأکید می‌کند که هیچ ابزار واحدی پاسخگوی تمامی نیازها نیست و کاربران باید بر اساس معیارهایی چون زبان متن، نوع تحلیل موردنظر (آماری، زبانی، آموزشی یا فنی)، حجم داده و هزینه، انتخاب آگاهانه‌ای داشته باشند.

در نهایت، یافته‌های این پژوهش نشان می‌دهد که اگرچه ابزارهای تحلیل متن رایانشی به‌سرعت در حال تکامل‌اند و توانمندی‌های چشمگیری پیدا کرده‌اند، اما همچنان به‌عنوان دستیاران هوشمند در کنار قضاوت انتقادی انسانی عمل می‌کنند و نمی‌توانند جایگزین تحلیل‌گر متخصص شوند. برای پژوهش‌های علمی قابل استناد و تحلیل‌های عمیق‌تر، استفاده از نرم‌افزارهای پیکره‌ای حرفه‌ای و کتابخانه‌های برنامه‌نویسی (مانند NLTK و spaCy برای فارسی) ضروری است. همچنین، توسعه ابزارهای بومی برای زبان فارسی، ایجاد مجموعه‌داده‌های استاندارد و ارتقاء سواد رایانشی کاربران از جمله گام‌های اساسی برای بهره‌برداری مؤثر از این فناوری‌ها در آینده خواهد بود.

منابع

- [۱] فلاحتی قدیمی فومنی، م. ر. و خسروانیان، م. (۱۴۰۲). نگاهی به چند نرم‌افزار کاربردی در حوزه تلفظ اسامی خاص. مجله‌ی علمی پژوهش در علوم رایانه، ۱(۲۹)، ۲۳-۳۴.
- [2] Alharbi, W. (2023). AI in the foreign language classroom: A pedagogical overview of automated writing assistance tools. *Education Research International*, 2023(1), 1-15. <https://doi.org/10.1155/2023/4253331>
- [3] Alsudais, A. (2025). Exploring the change in scientific readability following the release of ChatGPT. *Journal of Informetrics*, 19(3), Article 101679. <https://doi.org/10.1016/j.joi.2025.101679>
- [4] Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, 32, 1-16. <https://doi.org/10.1016/j.jslw.2016.01.003>
- [5] Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., & De Weerd, J. (2020). A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers & Education*, 162(3), Article 104094. <https://doi.org/10.1016/j.compedu.2020.104094>
- [6] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171-4186. <https://doi.org/10.18653/v1/N19-1423>
- [7] Ding, L., & Zou, D. (2024). Automated writing evaluation systems: A systematic review of Grammarly, Pigai, and Criterion with a perspective on future directions in the age of generative artificial intelligence. *Education and Information Technologies*, 29(11), 14151-14203. <https://doi.org/10.1007/s10639-023-12402-3>
- [8] Dizon, G., & Gayed, J. M. (2021). Examining the impact of Grammarly on the quality of mobile L2 writing. *The JALT CALL Journal*, 17(2), 74-92. <https://doi.org/10.29140/jaltcall.v17n2.336>
- [9] Fu, Q., Zou, D., Xie, H., & Cheng, G. (2022). A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning*, 37(2), 179-221. <https://doi.org/10.1080/09588221.2022.2033787>
- [10] Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Zh. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193-202. <https://doi.org/10.3758/BF03195564>
- [11] Janda, H. K., Pawar, A., Du, Sh., & Mago, V. (2019). Syntactic, semantic and sentiment analysis: The joint effect on automated essay evaluation. *IEEE Access*, 7, 108486-108503. <https://doi.org/10.1109/ACCESS.2019.2933354>
- [12] Ke, Z., & Ng, V. (2019). Automated essay scoring: A survey of the state of the art. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (pp. 6300-6308)*. IJCAI. <https://doi.org/10.24963/ijcai.2019/879>
- [13] Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, 96(2), 190-208. <https://doi.org/10.1111/j.1540-4781.2011.01232.x>
- [14] Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press.
- [15] Nunes, A., Cordeiro, C., Limpo, T., & Castro, S. L. (2021). Effectiveness of automated writing evaluation systems in school settings: A systematic review of studies from 2000 to 2020. *Journal of Computer Assisted Learning*, 38(2), 599-620. <https://doi.org/10.1111/jcal.12635>
- [16] Online Utility. Retrieved May 12, 2026, from <https://www.online-utility.org/text/analyzer.jsp>
- [17] Petchprasert, A. (2021). Utilizing an automated tool analysis to evaluate EFL students' writing performances. *Asian-Pacific Journal of Second and Foreign Language Education*, 6(1), 1-16. <https://doi.org/10.1186/s40862-020-00107-w>
- [18] Road To Grammar. Retrieved May 12, 2026, from <http://www.roadtogrammar.com/textanalysis/>
- [19] Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

- [20] Siregar, J., Nurlela, N., & Zein, T. T. (2023). An analysis of cohesive devices in EFL students' essay writing. *Indonesian Journal of EFL and Linguistics*, 8(1), 39-53. <https://doi.org/10.21462/ijefl.v8i1.577>
- [21] Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91(1), 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- [22] Takaki, P., & Dutra, M. L. (2024). Text mining applied to distance higher education: A systematic literature review. *Education and Information Technologies*, 29(9), 10851-10878. <https://doi.org/10.1007/s10639-023-12235-0>
- [23] Textalyser. Retrieved May 12, 2026, from <http://textalyser.net/>
- [24] Text Analysis Online. Retrieved May 12, 2026, from <http://textanalysisonline.com/>
- [25] Text fixer. Retrieved May 12, 2026, from <https://www.textfixer.com>
- [26] Usingenglish. Retrieved May 12, 2026, from <https://www.usingenglish.com/resources/text-statistics.php>
- [27] Wulandari, R. S., & Harida, R. (2021). Grammatical error analysis in essay writing. *DEIKSIS*, 13(1), 73-81. <https://doi.org/10.30998/deiksis.v13i1.5356>
- [28] Zhou, T., Cao, S., Zhou, S., Zhang, Y., & He, A. (2023). Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing. *System*, 118, Article 103141. <https://doi.org/10.1016/j.system.2023.103141>
- [29] Zhu, H., Wang, T., & Pang, N. (2024). Diachronic changes in lexical density of research article abstracts: A corpus-based study. *Lingua*, 312(3), Article 103837. <https://doi.org/10.1016/j.lingua.2024.103837>