



Analyzing the Application of Hyland Metadiscourse Model for Citation-based Automatic Text Summarization: A proposed Annotation Scheme for Citation Contexts

Pegah Tajer

Ph.D. Candidate in Knowledge and Information Retrieval, Department of Knowledge and Information Science, Shiraz University, Shiraz, Iran. Email: ptajer@shirazu.ac.ir

Abdorasoul Jowkar

*Corresponding Author, Professor, Department of Knowledge and Information Science, Shiraz University, Shiraz, Iran. Email: ajowkar2003@yahoo.com

Seyed Mostafa Fakhrahmad

Assistant Professor, Department of Computer Science and Engineering and Information Technology, Shiraz University, Shiraz, Iran. Email: fakhrahmad@shirazu.ac.ir

Hajar Sotoudeh

Associate Professor, Department of Knowledge and Information Science, Shiraz University, Shiraz, Iran. Email: sotoudeh@shirazu.ac.ir

Alireza Khormae

Associate Professor, Department of Foreign Languages and Linguistics, Shiraz University, Shiraz, Iran. Email: khoalireza23@gmail.com

Abstract

Objective: Author's abstract contains those contributions that the author himself considers important. Meanwhile, they may be less important among scientific community. This supplementary information can be obtained by analyzing citing articles. Citation contexts citing a cited article are actually summaries of that article produced by the scientific community. This type of summary is called citation summary which can provide a deeper insight into the impact of that article on scientific community. Selecting useful citation sentences to be inserted in a system summary is one of the major challenges of citation-based automatic text summarization. Hence, the semantic approach of analyzing citation contexts reveals citation functions; it can be used to refine citation contexts and to insert important content in the final summary. So, approaches like metadiscourse analysis that provide more information would result in producing

useful summaries. Therefore, this paper aims at analyzing the application of Hyland metadiscourse model for citation-based automatic summarization of scientific texts. Moreover, based on Hyland Metadiscourse Model, an annotation scheme was proposed for citation contexts which could be used in corpus-based citation summarization systems.

Methodology: This is a library research that answers research questions through studying and analyzing resources related to Hyland Metadiscourse Model, Scientific Text Summarization, Citation Context Analysis and Citation Function Classification. The scheme was evolved during two stages of analysis. First, an initial scheme was created based on studying existing schemes. Then, its metadiscourse version was suggested through analyzing Hyland Metadiscourse Model. Expert evaluation was performed for validating the proposed annotation scheme. Three experts in Information Science and two in Linguistics confirmed the scheme.

Findings: Hyland interactional metadiscourse is suitable for analyzing citation contexts because it is used to represent the author's perspective on propositional information and also the reader. Moreover, interactional metadiscourse analysis applies appropriate language tools for the critique genre. Therefore, a scheme was proposed based on boosters, attitude markers, hedges, engagement markers and self-mentions which are the main components of Hyland interactional metadiscourse. The proposed scheme includes 70 classes.

Conclusion: Hyland interactive metadiscourse can be used to construct proper corpora for automatic citation-based text summarization. Also, some other phases of automatic summarization such as classifier development, citation context refinement, and sentence selection could be performed based on this type of metadiscourse. Annotating corpora is usually performed using an annotation scheme. Thus, the proposed annotation scheme would be beneficial. However, it is a conceptual scheme proposed on existing theories. So, it is necessary to ask annotators to write down any new labels while annotating. Moreover, they should make some notes about the reasons of creating new ones. In the next stage, if desirable agreement is reached those labels could be added to the scheme.

Keywords: Annotation Scheme, Citation-based Summarization, Citation Contexts, Hyland Metadiscourse Model, Scientific Texts.

تحلیل کاربرد الگوی فراگفتمان هایلند در خلاصه‌سازی خودکار استنادمدار: پیشنهاد طرح حاشیه‌نویسی بافتارهای استنادی

پگاه تاجر

دانشجوی دکتری علم اطلاعات و دانش‌شناسی، گرایش بازیابی اطلاعات و دانش. دانشگاه شیراز، شیراز، ایران. رایانامه: ptajer@shirazu.ac.ir

عبدالرسول جوکار

*نویسنده مسئول، استاد گروه علم اطلاعات و دانش‌شناسی دانشگاه شیراز، دانشگاه شیراز، شیراز، ایران. رایانامه: ajowkar2003@yahoo.com

سید مصطفی فخر احمد

استادیار گروه مهندسی و علوم کامپیوتر و فناوری اطلاعات دانشگاه شیراز، شیراز، ایران. رایانامه: fakhrahmad@shirazu.ac.ir

هاجر ستوده

دانشیار گروه علم اطلاعات و دانش‌شناسی دانشگاه شیراز، شیراز، ایران. رایانامه: sotudeh@shirazu.ac.ir

علیرضا خرمایی

دانشیار گروه زبان‌های خارجی و زبان‌شناسی دانشگاه شیراز، شیراز، ایران. رایانامه: khoalireza23@gmail.com

چکیده

هدف: هدف مقاله حاضر، تحلیل کاربرد الگوی فراگفتمان هایلند در خلاصه‌سازی خودکار استنادمدار متون علمی و پیشنهاد یک طرح حاشیه‌نویسی فراگفتمان‌مدار برای بافتارهای استنادی به منظور به کارگیری در خلاصه‌سازی استنادمدار می‌باشد.

روش‌شناسی: روش‌شناسی این پژوهش از نوع کتابخانه‌ای است و پاسخ دهی به سؤالات پژوهش، از طریق مطالعه و تحلیل منابع مربوط به الگوی فراگفتمان هایلند، خلاصه‌سازی خودکار متون علمی، تحلیل بافتارهای استناد و طبقه‌بندی کارکردهای استنادی انجام شده است.

یافته‌ها: فراگفتمان تعاملی هایلند برای نشان دادن چشم‌انداز نویسنده نسبت به اطلاعات گزاره‌ای و خواننده به کار می‌رود، از ابزارهای زبانی مناسب ژانر نقد بهره می‌برد و برای تحلیل بافتارهای استنادی مناسب است. بنابراین، طرح حاشیه‌نویسی فراگفتمان‌مدار بافتارهای استنادی بر اساس تردیدنما، یقین‌نما، نگرش‌نما، خوداظهاریها و دخیل‌سازها که از مؤلفه‌های اصلی فراگفتمان تعاملی - مشارکتی هایلند هستند، پیشنهاد شد. این طرح شامل ۷۰ طبقه می‌باشد.

نتیجه‌گیری: از فراگفتمان تعاملی هایلند می‌توان برای ساخت پیکره مناسب جهت خلاصه‌سازی خودکار استنادمدار بهره گرفت و مراحل ایجاد رده‌بندهای مورد نیاز فرآیند خلاصه‌سازی، پالایش بافتارهای استنادی و انتخاب جملات برای درج در خلاصه نهایی را بر اساس آن انجام داد. حاشیه‌نویسی پیکره‌ها عموماً بر اساس یک طرح حاشیه‌نویسی انجام می‌شود. بنابراین، طرح پیشنهاد شده می‌تواند مفید واقع شود. با توجه به این که طرح حاشیه‌نویسی پیشنهاد شده مبتنی بر نظریات موجود است، لازم است در به کارگیری آن، از حاشیه‌نویسان خواسته شود تا در حین برچسب‌زنی، هر برچسب دیگری غیر از موارد مطرح شده در طرح را که به ذهنشان می‌رسد با ذکر دلیل، یادداشت نمایند تا در صورت احراز توافق مطلوب به طرح اضافه گردد.

کلیدواژه‌ها: فراگفتمان هایلند، بافتارهای استنادی، خلاصه‌سازی استنادمدار، طرح حاشیه‌نویسی، متون علمی.

کتابخانه مرکزی آستان قدس رضوی

کتب‌داری و اطلاع‌رسانی، دوره ۲۲، شماره ۳، پاییز ۱۳۹۸، صص. ۹۱-۱۱۱.

تاریخ ارسال: ۹۷/۸/۱۵ - تاریخ پذیرش: ۹۷/۱۱/۲

مقدمه

مقالات علمی عموماً دارای چکیده نویسنده می‌باشند که هدف، روش، یافته‌ها و نتیجه‌گیری نویسنده را به اختصار ارائه می‌دهد. در این چکیده‌ها، سهم نویسنده در پیشبرد دانش به صورت عام بیان می‌شود. به عبارت دیگر، چکیده نویسنده شامل آن دسته از اصالت اثر است که نویسنده آن‌ها را مهم دانسته و در چکیده خود ذکر کرده است. این در حالی است که ممکن است موارد مورد نظر نویسنده، در میان جامعه علمی اهمیت کمتری داشته باشد. بنابراین، چکیده نویسنده در بر دارنده اطلاعاتی در مورد تأثیر مقاله در علم نیست. این اطلاعات تکمیلی را می‌توان با رجوع به مقالات استنادکننده به آن مقاله، به دست آورد. مجموعه محتوای استنادها به یک مقاله را در واقع می‌توان خلاصه‌ای از آن مقاله دانست که توسط جامعه علمی تولید شده است. این نوع خلاصه، خلاصه استنادی^۱ نام دارد (الکیس، شن، فیدر، ارکان، استیتز و رادو^۲، ۲۰۰۸؛ قزوینیان، رادو، محمد، دور، زاجیک، ویدبای و مون^۳، ۲۰۱۳؛ کهن، سولدینی^۴ و گوهریان، ۲۰۱۴). یک خلاصه استنادی می‌تواند تأثیر مقاله در جامعه علمی را آشکار سازد، همچنین می‌تواند در بردارنده مقایسه‌هایی با یافته‌های مشابه از مقالات دیگر باشد. به این ترتیب، یک خلاصه استنادی با فراهم آوردن اطلاعات اضافی نسبت به چکیده نویسنده، می‌تواند بینش عمیق‌تری راجع به تأثیر آن مقاله در جامعه علمی را فراهم آورد.

یکی از حوزه‌هایی که بهره‌گیری از محتوای جملات استنادی را مورد توجه قرار داده است، حوزه بازیابی اطلاعات علمی^۵ است. این امر باعث شده است تا در ادبیات خلاصه‌سازی خودکار متون، نوعی از خلاصه‌سازی با نام خلاصه‌سازی استنادمدار^۶ پدید آید. این نوع خلاصه‌سازی از مجموعه‌ای از استنادها به یک مقاله مرجع، برای تولید خلاصه بهره می‌برد (قزوینیان و رادو، ۲۰۰۸؛ قزوینیان و همکاران، ۲۰۱۳).

چگونگی انتخاب محتوای مفید جملات استناددهنده برای درج در خلاصه نهایی، از چالش‌های مهم پژوهش‌های خلاصه‌سازی خودکار^۷ استنادمدار است. از آنجایی که رویکرد معنایی تحلیل بافتارهای استناد^۸، کارکردهای استناددهی^۹ را آشکار می‌کند، می‌توان از آن برای پالایش محتوای بافتارها و درج محتوای مهم در

1. Citation Summary

2. Elkiss, Shen, Fader, Erkan, States & Radev

3. Dorr, Zajic, Whidby & Moon

4. Soldaini

5. Literature Retrieval

6. Citation-based summarization

۷. خلاصه‌سازی خودکار متن از مباحث مهم و پرچالش حوزه پردازش زبان طبیعی است و عبارت است از نمایش فشرده، دقیق و منسجم متن ورودی به طوری که متن خروجی، مفاهیم مهم متن ورودی را در برداشته باشد.

۸. تحلیل استنادی محتوا محور، در دو سطح نحوی و معنایی انجام می‌شود. در سطح نحوی می‌توان با ارزش‌گذاری به جایگاه استناد، بسامد متنی استناد و ویژگی‌های ساختاری مقالات، استنادها را متمایز ساخت. سطح بالاتر، سطح معنایی است که می‌توان استنادها را بر اساس سرشت آن‌ها مانند نوع استناد (تطبیقی، استناد به روش‌شناسی و غیره) تحلیل نمود (قدیمی و ستوده، ۱۳۹۵)

9. Citation Functions

خلاصه استفاده کرد. با این وجود، اگر در تحلیل بافتارهای استنادی، از رویکردهایی که اطلاعات بیشتری در اختیار قرار می‌دهند و به عبارت دیگر، مطالعه زبان را در سطح بالاتر از جمله امکان‌پذیر می‌سازند؛ استفاده شود می‌توان به پالایش عمیق‌تر آن‌ها و درج بافتارهای مفیدتر در خلاصه‌های خودکار نهایی، امیدوارتر بود.

یکی از این رویکردها، تجزیه و تحلیل کلام است. فراگفتمان^۱ یکی از مباحث مهم تجزیه و تحلیل کلام می‌باشد که در واقع حلقه ارتباط بین متن، نویسنده و خواننده است (هایلند و تسه^۲، ۲۰۰۴؛ هایلند، ۲۰۰۵ الف). فراگفتمان به نویسنده یا گوینده این امکان را می‌دهد که نظرات خود را بیان کند و همچنین با مخاطب به منزله عضوی از یک جامعه خاص ارتباط برقرار نماید (هایلند، ۲۰۰۵ الف). به عبارت دیگر، فراگفتمان بیان این ایده است که ارتباط زبانی فراتر از تبادل صرف اطلاعات و گزاره‌هاست و در حقیقت شامل هویت‌ها، نگرش‌ها و فرض‌های طرفین ارتباط است (هایلند، ۲۰۰۵ الف نقل در طارمی و دیگران، ۱۳۹۷). بنا بر آنچه گفته شد، به نظر می‌رسد که بهره‌گیری از آن در پالایش انواع بافتارهای استنادی و استخراج و انتخاب بافتارهای حاوی نقد و نگرش نویسنده استناددهنده راهگشا باشد.

تحلیل فراگفتمان با بهره‌گیری از نشانگرهای فراگفتمانی انجام می‌شود. نشانگرهای فراگفتمان، در واقع صورت‌های زبانی هستند که از ابزارهای مهم بلاغی ایجاد و تضمین کننده تعامل در متون محسوب می‌شوند (طارمی و دیگران، ۱۳۹۷). الگوهای مختلفی برای تحلیل فراگفتمان ارائه شده است. یکی از جامع‌ترین و مشهورترین این الگوها، الگوی هایلند (۲۰۰۵ ب) می‌باشد. وی، مطالعات گسترده‌ای در زانر علم انجام داده است و الگوی فراگفتمان خود را مبتنی بر تحلیل نوشتارهای دانشگاهی توسعه داده است (هایلند، ۱۹۹۸؛ هایلند و تسه، ۲۰۰۴؛ هایلند، ۲۰۰۴). پژوهش‌های وی، بیانگر این مهم است که متن علمی پدیده‌ای خنثی، عینی و ایستا نیست بلکه ضمن هدایت دیدگاه خواننده، دنیای ذهنی شامل افکار، تصورات، نگرش‌ها و بعضاً ویژگی‌های فردی و اجتماعی نویسنده متن را بازتاب می‌دهد. لازم به ذکر است که تاکنون، از رویکرد تحلیل فراگفتمان برای خلاصه‌سازی استنادمدار استفاده نشده است.

با توجه به مطالب گفته شده در بالا، این پژوهش، با هدف تحلیل نظری کاربرد الگوی فراگفتمان هایلند در خلاصه‌سازی استنادمدار و پیشنهاد یک طرح حاشیه‌نویسی فراگفتمان‌مدار برای بافتارهای استنادی که از آن بتوان در خلاصه‌سازی استنادمدار مبتنی بر پیکره بهره‌گرفت، درصدد است به پرسش‌های زیر پاسخ دهد:

- از منظر نظر، آیا فراگفتمان‌های موجود در الگوی هایلند، برای حاشیه‌نویسی بافتارهای استنادی مورد

استفاده در خلاصه‌سازهای خودکار استنادمدار مناسب می‌باشد؟

1. Metadiscourse

2. Hyland & Tse

- از منظر نظر، یک طرح حاشیه‌نویسی فراگفتمان‌مدار بافتارهای استنادی چگونه است؟

پیشینه پژوهش

مطالعات تویفل و مونس در سال ۲۰۰۲ را می‌توان اولین پژوهش خلاصه‌سازی خودکار ژانر علمی در ادبیات پردازش زبان طبیعی دانست (تویفل و مونس^۱، ۲۰۰۲) اما ریشه‌های خلاصه‌سازی خودکار متون علمی را می‌توان در ادبیات علم اطلاعات و دانش‌شناسی و مباحث چکیده‌نویسی خودکار سال‌های دهه پنجاه قرن بیستم جست (لون^۲، ۱۹۵۸؛ باکسندیل^۳، ۱۹۵۸). در واقع، تمایل به خلاصه‌سازی خودکار متون علمی از زمانی آغاز شد که کتاب‌ها و مقالات علمی به صورت دیجیتالی ذخیره شدند. با این وجود، جامعه پژوهشی خلاصه‌سازی خودکار، از خلاصه‌سازی مقالات علمی به خلاصه‌سازی مقالات خبری روی آورد و این روند تا دوباره فعال شدن خلاصه‌سازی خودکار ژانر علمی در اواخر دهه نود میلادی، همچنان ادامه داشت (یلوگلو، میلیوس و زینکر-هیوود^۴، ۲۰۱۱).

اغلب پژوهش‌های خلاصه‌سازی خودکار در ژانر علمی از روش‌های یادگیری ماشینی بهره می‌برند. الگوریتم‌های خلاصه‌ساز ارائه شده توسط کوپیک، پدرسن و چن^۵ (۱۹۹۵)، تویفل و مونس (۲۰۰۲) مبتنی بر یادگیری ماشینی نظارتی و برخی دیگر مانند الگوریتم قزوینیان و رادو (۲۰۰۸ و ۲۰۱۰)، الکیس و همکاران (۲۰۰۸) و ابوجبارا^۶ و رادو (۲۰۱۱) بر رویکرد غیرنظارتی بنا نهاده شده‌اند. الگوریتم‌های خلاصه‌سازی ارائه شده، عموماً تلفیقی از روش‌های سطحی و معنایی می‌باشند. دو استراتژی معنایی مشهور در خلاصه‌سازهای نظارتی متون علمی عبارتند از: ناحیه‌بندی استدلالی^۷ و مفاهیم هسته^۸.

در ناحیه‌بندی استدلالی، فرض بر این است که گفتمان علمی در بردارنده اظهارنظرهای مختلف درباره سهم پژوهشگران از یک حوزه علمی است. این نوع گفتمان، حاصل یک بازی بلاغی است که سعی دارد جایگاه و سهم پژوهشگران در پیشبرد دانش را ارتقاء دهد. بر این اساس، متون علمی باید ادعاهای دانش را در تقابل با آثار پژوهشی قبلی به وضوح نشان دهند (فیسس، رونزانو و ساگیون^۹، ۲۰۱۵).

1. Teufel & Moens
2. Luhn
3. Baxendale
4. Yeloglu, Milios & Zincir-Heywood
5. Kupiec, Pedersen & Chen
6. Abu-Jbara
7. Argumentative Zoning (AZ)
8. Core Scientific Concepts (CoreSc)
9. Fisas, Ronzano & Saggion

از این منظر، الگوی گفتمانی ادعاهای دانش^۱ معروف به الگوی سیمون تویفل^۲ یا همان ناحیه‌بندی استدلالی به وجود آمده است (تویفل، ۱۹۹۹؛ تویفل و مونس، ۲۰۰۲؛ تویفل، سیدهارتن و بچلور^۳، ۲۰۰۹). در رویکرد خلاصه‌سازی مبتنی بر ناحیه‌بندی استدلالی، وضعیت بلاغی هر جمله به منظور بازنمون بافت گفتمانی جملات استخراج شده تعیین می‌شود؛ به عبارت دیگر، در این استراتژی، فرآیند استخراج جمله با تحلیل گفتمان ترکیب شده است.

در استراتژی مفاهیم هسته، فرض بر این است که یک مقاله علمی، بازنمونی از یک بررسی علمی قابل خواندن توسط انسان است که با بررسی آن می‌توان، به چگونگی مؤلفه‌های پژوهشی و محل قرارگیری آن‌ها در متن پی برد. بر اساس این استراتژی، جملات مقاله علمی با توجه به مفاهیمی از هستان‌شناسی^۴ بررسی‌های علمی، حاشیه‌نویسی می‌شوند. این نوع هستان‌شناسی، مجموعه مفاهیم لازم برای توصیف یک بررسی علمی را در بر دارد.

استراتژی مفاهیم هسته، برای طبقه‌بندی‌های محتوامدار یک بررسی علمی و استراتژی ناحیه‌بندی بلاغی، برای طبقه‌بندی‌های مبتنی بر ادعاهای دانش با رویکرد استنادی مفید می‌باشد (فیسس، رونزانو و ساگیون، ۲۰۱۵؛ لیاکاتا^۵، تویفل، سیدهارتان و بچلور، ۲۰۱۰).

در هر دو استراتژی، بعد از تحلیل متن (تعیین ناحیه استدلالی و یا استخراج مفاهیم هسته)، می‌توان برای طبقه‌بندی و انتخاب جملات مهم به منظور درج در خلاصه نهایی اقدام نمود. لازم به ذکر است که استراتژی‌های موجود منجر به ظهور طرح‌های حاشیه‌نویسی^۶ مبتنی بر تحلیل معنایی در سطح کلمه، جمله و بعضاً در سطح گفتمان علمی شده‌اند^۷ (تویفل، سیدهارتان و تیدهار^۸، ۲۰۰۶؛ یورگنس، کومار، هوور، مک‌فارلند و جورافسکی^۹، ۲۰۱۶؛ هرماندز-آلوارز و گومز^{۱۰}، ۲۰۱۶).

1. Knowledge Claim Discourse Model (KCDM)

2. Simone Teufel's model

3. Siddharthan & Batchelor

4. Ontology

5. Liakata

6. Annotation schemes

۷. برای مطالعه طرح‌های حاشیه‌نویسی متون علمی رجوع شود به:

تاجر، پگاه؛ جوکار، عبدالرسول؛ فخراحمد، سیدمصطفی؛ خرماپی، علیرضا؛ ستوده، هاجر (۱۳۹۶). کاربرد تحلیل گفتمان در خلاصه‌سازی خودکار متون علمی. مجموعه مقالات نخستین همایش ملی رویکردهای نوین در مطالعات زبان و ادبیات، ۲۶ بهمن، مؤسسه آموزش عالی زند، شیراز، ایران.

8. Tidhar

9. Jurgens, Kumar, Hoover, McFarland & Jurafsky

10. Hernández-Alvarez & Gómez

مرور پیشینه پژوهش نشان داد که تاکنون پژوهشی که از الگوهای فراگفتمانی موجود برای ارائه طرح‌های حاشیه‌نویسی بافتارهای استنادی به هدف استفاده در سامانه‌های خلاصه‌ساز استنادمدار بهره ببرد، انجام نشده است.

از سوی دیگر، بررسی نوشتارها نشان داد که الگوی فراگفتمانی هایلند، به طور وسیعی از طرف جامعه علمی، پذیرفته شده است (طارمی، تاکی و یوسفیان، ۱۳۹۷). لذا، در این پژوهش به کار گرفته شد. این الگو، پایه نظری این پژوهش را تشکیل می‌دهد که در ادامه تشریح شده است.

مبانی نظری پژوهش

هایلند، برای نخستین بار، تمام جنبه‌های مربوط به تعامل میان نویسنده متن و خود متن و همچنین نویسنده متن و خواننده متن را ذیل اصطلاح «فراگفتمان» تعریف و طبقه‌بندی کرده و ادعا می‌کند محقق در چارچوب مطالعه مبتنی بر فراگفتمان می‌تواند از دایره متن و گفتمان فراتر رفته و متن را به مثابه کنشی اجتماعی تحلیل کند (قلی مافیان و کارگر، ۱۳۹۳).

از نظر وی، نویسنده متن، آگاهانه یا ناآگاهانه از ابزارهای مختلف زبانی برای تعامل با خواننده استفاده می‌کند. هایلند، فراگفتمان تعاملی - هدایتی^۱ و فراگفتمان تعاملی - مشارکتی^۲ را بیان کرد. فراگفتمان تعاملی - هدایتی با سازماندهی اطلاعات گزاره‌ای به نحوی عمل می‌کند که متن را منسجم و برای مخاطب متقاعدکننده جلوه دهد. این در حالی است که فراگفتمان تعاملی - مشارکتی، برای نشان دادن چشم‌انداز نویسنده نسبت به اطلاعات گزاره‌ای و خواننده به کار می‌رود (هایلند، ۲۰۰۵ الف نقل در طارمی و دیگران، ۱۳۹۷).

هایلند (۲۰۰۵ ب) در الگوی خود، ۱۰ نشانگر فراگفتمانی را ارائه می‌نماید که پنج نشانگر مربوط به فراگفتمان تعاملی - هدایتی و پنج نشانگر به فراگفتمان تعاملی - مشارکتی مرتبط هستند. نشانگرهای فراگفتمانی تعاملی - هدایتی عبارتند از: گذارها^۳، قالب‌نماها^۴، ارجاعات^۵ درون متنی^۶، استنادها^۷ و ابهام‌زداها^۸.

1. Interactive
2. Interactional
3. Transitions
4. Frame Markers

۵. در زبان‌شناسی، به حالتی در جمله‌ها که درک معنای یک عنصر متنی با مراجعه به عناصر دیگر متن امکان‌پذیر می‌شود ارجاع گویند. عمده ارجاعات را ضمائر شخصی، ملکی، اشاره‌ای و صفات ملکی تشکیل می‌دهند.

6. Endophoric Markers
7. Evidentials
8. Code Glosses

نشانه‌های فراگفتمان تعاملی- مشارکتی عبارتند از: تردیدنما، یقیننما، نگرش‌نما، خود اظهار^۴ و دخیل‌ساز^۵. جدول ۱ توضیح هر یک از نشانه‌ها را به همراه مثال نشان می‌دهد.

جدول ۱. طبقه‌بندی نشانه‌های فراگفتمانی هایلند (هایلند، ۲۰۰۵ الف)

نشانه	کاربرد	مثال
تعاملی - هدایتی	راهنمایی خواننده در متن	منابع^۶
گذارها	روابط بین جمله‌واره‌های اصلی ^۷ را بیان می‌کنند.	in addition; but; thus; and
قالب نماها	به مراحل، توالی‌ها و «اعمال گفتمانی» ^۸ اشاره می‌کنند.	finally; to conclude; my purpose is
ارجاعات درون متنی	به اطلاعات ارائه شده در قسمت‌های دیگر متن اشاره می‌کنند.	noted above; see Fig; in section 2
استنادها (شواهد)	به اطلاعات ارائه شده در متون دیگر اشاره می‌کنند.	according to X; Z states
ابهام زداهای	معانی محتوای گزاره‌ای را با دقت شرح می‌دهند ^۹ .	namely; e.g.; such as; in other words
تعاملی - مشارکتی	به مشارکت طلبیدن خواننده در متن	منابع
تردیدنماها	عبارت‌های احتیاطی هستند که از هرگونه التزام خودداری می‌کنند و گفتگو را باز می‌کنند.	might; perhaps; possible; about
یقین‌نماها	عبارت‌های تأکیدی هستند که قطعیت و یقین نویسنده را نسبت به تحقق امری بازتاب می‌دهند و گفتگو را می‌بندند.	in fact; definitely; it is clear that
نگرش‌نماها	دیدگاه و نگرش نویسنده را در خصوص یک پدیده بیان می‌کنند و نیز بیان‌کننده شگفتی، اجبار، موافقت؛ اهمیت و غیره است.	unfortunately; I agree; surprisingly
خود اظهارها	نویسنده با استفاده از ضمائر شخصی و ملکی در گفتمان، به طور صریح به خودش ارجاع می‌دهد.	I; we; my; me; our
دخیل‌سازها	برای ارتباط نویسنده با مخاطب و به عبارتی مشارکت جویی نویسنده از طریق جلب توجه آن‌ها به موضوع مورد بحث یا شرکت دادن آن‌ها در گفتمان از طریق ضمائر دوم شخص، پاره‌گفت‌های امری، پرسش و غیره کاربرد دارد.	consider; note; you can see that

1. Hedges
2. Boosters
3. Attitude Markers
4. Self-mentions
5. Engagement Markers
6. Resources
7. Main clauses
8. Discourse acts
9. Elaborate

روش پژوهش

در این پژوهش، از روش کتابخانه‌ای استفاده شده است. پژوهشگر با استفاده از جستجوی اینترنتی به متون مربوط به الگوی فراگفتمان هایلند، خلاصه‌سازی خودکار متون علمی، تحلیل بافتارهای استناد و طبقه‌بندی کارکردهای استناددهی دست یافت. سپس، با مطالعه منابع، به تحلیل کاربرد الگوی فراگفتمان هایلند در خلاصه‌سازی خودکار پرداخت و یک طرح حاشیه‌نویسی فراگفتمان‌مدار برای بافتارهای استنادی پیشنهاد داد. این طرح، در طی دو مرحله تحلیل ارائه گردید. ابتدا بر اساس طرح‌های موجود، طرح اولیه‌ای ایجاد شد. سپس، نسخه فراگفتمان‌مدار طرح بر اساس الگوی هایلند ارائه گردید. لازم به ذکر است که به منظور اعتباریابی، طرح حاشیه‌نویسی پیشنهاد شده در اختیار متخصصان شامل ۳ نفر متخصص حوزه علم اطلاعات و ۲ نفر متخصص حوزه زبان‌شناسی قرار گرفت.

یافته‌های پژوهش

از منظر نظر، آیا فراگفتمان‌های موجود در الگوی هایلند، برای حاشیه‌نویسی بافتارهای استنادی مورد استفاده در خلاصه‌سازهای خودکار استنادمدار مناسب هستند؟

فراگفتمان تعاملی - هدایتی هایلند (۲۰۰۵ ب)، به آگاهی نویسنده از مخاطب، روش‌هایی برای قرار دادن دانش، علاقه‌مندی‌ها، عبارات بلاغی در متن و توانایی‌های پردازش نویسنده مربوط است. هدف نویسنده در این بعد، شکل دادن به متن به صورتی است که نیازهای خوانندگان خاص برآورده گردد. علاوه بر این، نویسنده در این بعد تلاش می‌کند، استدلال‌ات خود را طوری سازماندهی نماید تا خواننده بتواند آسان‌تر، متن را دنبال کند (هایلند، ۲۰۰۵ الف. نقل در طارمی و دیگران، ۱۳۹۷). بنابراین، به کارگیری منابع زبانی فراگفتاری در این بعد، بیشتر به منظور سازماندهی گفتمان است و حد و اندازه‌ای که متن، بر اساس نیازهای خوانندگان موجود در ذهن نویسنده ساخته شده است را آشکار می‌کند. به عبارت دیگر، این نوع فراگفتمان، با به کارگیری مجموعه‌ای از امکانات زبانی مانند گذارها، چارچوب‌سازها و غیره، به نویسنده یا گوینده این امکان را می‌دهد که اطلاعات گزاره‌ای را مدیریت کند؛ بنابراین، به نظر نمی‌رسد که این نوع از فراگفتمان را بتوان برای تحلیل و پالایش بافتارهای استنادی به کار برد و لازم است از مؤلفه‌هایی که می‌توانند بینش عمیق‌تری از محتوای بافتارها را ارائه کنند، استفاده کرد.

فراگفتمان تعاملی - مشارکتی هایلند (۲۰۰۵ ب) به روش‌هایی مربوط است که نویسندگان از طریق اظهارنظر بر پیامشان سعی به برقراری تعامل دارند. در این بعد، هدف نویسنده بیان کردن آشکار دیدگاه‌هایش و به مشارکت طلبیدن خوانندگان از طریق اجازه دادن به آن‌ها برای پاسخ‌دهی به متن است. فراگفتار تعاملی -

مشارکتی، اساساً ارزش‌گذارانه و مشارکت‌طلبانه است و در تلاش است همکاری نویسنده و خواننده را نشان دهد و اعتراضات را پیش‌بینی نماید. این بعد آشکار می‌کند که تا چه حد، نویسنده تلاش کرده است تا متن را با به مشارکت طلبیدن خوانندگان، بسازد (هایلند، ۲۰۰۵ الف). به عبارت دیگر، فراگفتمان تعاملی - مشارکتی هایلند (۲۰۰۵ ب)، برای نشان دادن چشم‌انداز نویسنده نسبت به اطلاعات گزاره‌ای و خواننده به کار می‌رود. در این نوع فراگفتمان، نویسنده یا گوینده از ابزارهای زبانی مانند یقین‌نماها، تردیدنماها و غیره به منظور ابراز نگرش خود و جلب مشارکت خواننده استفاده می‌کند. از آنجایی که نقش این ابزارهای زبانی در ژانر نقد مهم‌تر هستند (قلی مافیان و کارگر، ۱۳۹۳)، به‌کارگیری فراگفتمان تعاملی - مشارکتی برای ایجاد پیکره‌های حاشیه‌نویسی شده بافتارهای استنادی مناسب‌تر می‌نماید؛ زیرا به کمک این ابزارها می‌توان به درک عمیق‌تر بافتارهای استنادی در بردارنده محتوای نقد و تحلیل ادعاهای دانش امیدوارتر بود.

از منظر نظر، یک طرح حاشیه‌نویسی فراگفتمان‌مدار بافتارهای استنادی چگونه است؟

مرور نوشتارها نشان داد که طرح‌های حاشیه‌نویسی موجود متون علمی، عموماً از روش‌های معنایی و گفتمانی بهره می‌برند که عمدتاً مبتنی بر نظریه ساختار بلاغی^۱ متن می‌باشند. از آنجایی که پژوهشگران درصدد به کارگیری رویکرد فراگفتمان در خلاصه‌سازی خودکار استنادمدار هستند، لازم است طرحی برای حاشیه‌نویسی فراگفتمان‌مدار بافتارهای استنادی ارائه نمایند. همان‌طور که در روش‌شناسی گفته شد، این طرح طی دو مرحله تحلیل پیشنهاد شد. در ادامه، این دو مرحله تشریح شده‌اند.

مرحله ۱: شکل‌گیری طرح حاشیه‌نویسی اولیه

بررسی متون منتشر شده نشان داد که سیاهه‌های متنوعی از دلایل و انگیزه‌های استناددهی و کارکردهای آن توسط پژوهشگران پیشنهاد شده است که اشتراکاتی نیز با یکدیگر دارند. یکی از جامع‌ترین سیاهه‌های انگیزه‌های استناددهی توسط گارفیلد^۲ (۱۹۶۵) ارائه شده است.

از طرف دیگر، مطالعه ادبیات تحلیل بافتار استناد نشان داد که حاشیه‌نویسی مجموعه داده‌ها، یکی از روش‌های تحلیل بافتار استناد می‌باشد (تویفل، ۱۹۹۹؛ تویفل، ۲۰۱۰؛ اطهر^۳، ۲۰۱۴؛ هرناندز-آلوارز و گومز،

1. Rhetorical Structure Theory (RST)

این نظریه، یکی از نظریه‌های مشهور سازماندهی گفتمان می‌باشد که بر اساس آن می‌توان متن را به صورت ساختاری درختی یا درختی بلاغی بازنمون نمود که در آن محدوده‌های متنی^۱ با به کارگیری مجموعه‌ای از روابط گفتمانی از پیش تعیین شده به هم پیوند داده می‌شوند. با توجه به این که متن یک واحد زبان‌شناسی پیچیده است، ساختار گفتمانی و یا تئوری‌های سازماندهی متن را می‌توان شاخصی برای تفسیر متن در نظر گرفت.

2. Garfield

3. Athar

۲۰۱۶). بهره‌گیری از این روش، منجر به ایجاد طرح‌های حاشیه‌نویسی کارکردهای استناد شده است اما هیچ کدام از آن‌ها فراگفتمان‌مدار نیستند. لازم به ذکر است که در طبقه‌بندی کارکردهای استنادی از دسته‌بندی‌ها و طرح‌های گوناگونی استفاده شده است و توافق جمعی بر روی یک طرح طبقه‌بندی واحد وجود ندارد؛ مثلاً، تویفل (۱۹۹۹) کارکردهای استنادی را در هفت دسته اطلاعات زمینه‌ای، مالکیت^۱، هدف، متنی، تضاد، پایه و سایر قرار داد، حال آن که لیاکاتا، ساها^۲، دوبنیک^۳، بچلور و ربولز-شومان^۴ (۲۰۱۲) طرح رده‌بندی کارکرد استنادی خود را با ۱۱ طبقه ارائه دادند که عبارتند از: فرضیه، انگیزه، رسالت، هدف، پیشینه، روش، آزمایش، مدل، مشاهده، یافته و نتیجه.

با توجه به نظر پژوهشگران و اساتید علم‌سنجی، مبتنی بر جامع‌تر بودن الگوی انگیزه‌های استناددهی گارفیلد (۱۹۶۵) نسبت به دیگر الگوهای موجود، سیاهه وی، مبنای طرح اولیه حاشیه‌نویسی قرار گرفت. لازم به ذکر است که آن دسته از دلایل استناددهی گارفیلد (۱۹۶۵) که به نظر می‌رسد با تحلیل محتوای گفتامانی بافتار استناد قابل شناسایی باشند، به عنوان پایه‌ای برای ایجاد طرح حاشیه‌نویسی در نظر گرفته شدند؛ زیرا درک تعدادی از انگیزه‌های استناددهی ارائه شده در الگوی گارفیلد، بسیار ذهنی هستند و به تحلیل نویسنده و مصاحبه با وی نیاز دارند. انگیزه‌هایی مانند «ادای احترام و قدرشناسی نسبت به متقدمان»، «فراهم‌آوردن راهنمایی از آثار دیگران» و «شناسایی انتشارات بنیادی درباره مفهومی خاص» از این نوع هستند. از آنجایی که هدف، ارائه طرحی برای حاشیه‌نویسی متن می‌باشد؛ انگیزه‌هایی که به تحلیل نویسنده نیاز دارند، لحاظ نشدند. بدین منظور، با اساتید علم‌سنجی نیز مشورت شد و موارد انتخابی را تأیید کردند. نهایتاً، ۹ مورد از انگیزه‌های استناددهی گارفیلد برای ایجاد طرح اولیه مورد توجه قرار گرفت. این ۹ مورد عبارتند از:

معرفی روش و ابزارهای علمی مورد استفاده، فراهم‌آوری زمینه برای مطالعه بیشتر، تصحیح اثر خود، تصحیح آثار دیگران، نقد آثار دیگران، اثبات یک ادعا، آگاه کردن محققان از آثار نشریافته و یا در حال انتشار، رد اثر یا اندیشه منتسب به دیگران و رد ادعاهای حق تقدم در آثار دیگران.

از طرف دیگر، با توجه به این که طرح‌های موجود طبقه‌بندی کارکردهای استناد و نوع آن‌ها، اشتراکاتی با یکدیگر دارند، این اشتراکات در طرح‌ریزی اولیه، در نظر گرفته شد. جدول ۲، طرح اولیه حاشیه‌نویسی بافتارهای استنادی را نشان می‌دهد. این طرح شامل ۱۵ طبقه است. طبقات، برآیندی از مرور طرح‌های

1. OWN
2. Saha
3. Dobnik
4. Rebholz-Schuhmann

موجود و اعمال تحلیل‌های نظری گفته شده در بالا می‌باشند. برای هر طبقه نیز نامی قراردادی تخصیص داده شد.

جدول ۲. طرح حاشیه‌نویسی اولیه

نام برچسب	توضیحات برچسب
BKG	اطلاعات زمینه‌ای
LIT-REW	مرور پیشینه (آگاه کردن محققان از آثار نشر یافته یا در حال انتشار، بیان نقاط قوت یا ضعف اثر مورد استناد/ ارزیابی اثر در بخش پیشینه پژوهش)
PBM	بیان مسئله پژوهش و اهمیت آن (شکاف دانشی و اهمیت و ضرورت بررسی آن)
CONTI	نویسنده به قصد ادامه و گسترش نتایج پژوهش قبلی خودش استناد داده است.
EXTEN	نویسنده به قصد توسعه داده‌ها، روش‌ها و رویکردهای دیگران استناد داده است.
FUT	بیان پژوهش‌های آینده
RSL-ANL	تحلیل یافته‌ها
CON-COMP-SUP	مقایسه نتایج پژوهشی - ارائه شواهد همسو
CON-COMP- DFF	مقایسه نتایج پژوهشی - غیرهمسو بودن نتایج: مغایرت‌ها و تناقض‌ها
MTH	معرفی روش و ابزارهای علمی مورد استفاده، توجیه چرایی انتخاب جامعه آماری و دیگر مباحث روش‌شناسی پژوهش
OWN-CRR	تصحیح اثر خود
OTHER-CRR	تصحیح آثار دیگران
PROV-CLM	اثبات یک ادعا
OTHER-REJ	رد اثر یا اندیشه متناسب به دیگران
PRIOR-REJ	رد ادعاهای حق تقدم در آثار دیگران

مرحله ۲: شکل‌گیری طرح حاشیه‌نویسی فراگفتمان‌مدار

در مرحله بعد، به منظور ارائه نسخه فراگفتمان‌مدار طرح حاشیه‌نویسی بافتارهای استنادی، لازم بود با توجه به تعاریف مفهومی ارائه شده از دو نوع فراگفتمان ارائه شده در الگوی هایلند، مؤلفه‌های مناسب به طرح اضافه گردد و سپس، طبقات بر اساس آن‌ها بازنگری شوند.

بدین ترتیب، مؤلفه‌های فراگفتمان تعاملی - مشارکتی هایلند (۲۰۰۵ ب) در طرح پیشنهادی حاشیه‌نویسی فراگفتمان‌مدار بافتارهای استنادی لحاظ شد. طرح پیشنهادی شامل ۷۰ طبقه می‌باشد. نگاه کنید به جدول ۳.

تحلیل فراگفتمان مبتنی بر نشانگرهای تردیدنما، یقین‌نما، نگرش‌نما و دخیل‌ساز بینشی از احتیاطی، تأکیدی، نگرشی و مشارکت‌جویانه بودن گفتمان را آشکار می‌سازند (خانی و چنگیزی، ۱۳۹۲). لذا، این الفاظ در نام‌گذاری طبقات به کار گرفته شد.

از آنجایی که به نظر می‌رسد در بافتارهای استناد، حضور خود اظهارها نشان‌های برای شناسایی خوداستنادی‌ها باشد، در طرح‌ریزی طبقات مورد توجه قرار گرفتند. لازم به ذکر است که طبقه «ادامه و گسترش نتایج پژوهش قبلی نویسنده» ذاتاً خوداستنادی است و در همان مرحله اول تحلیل، یعنی تحلیل محتوای گزاره‌ای متن، آشکار می‌شود. این در حالی است که ماهیت فراگفتمانی آن را می‌توان، با دیگر ابزارهای فراگفتمان تعاملی - مشارکتی مانند تردیدنماها و غیره معلوم نمود که در طرح‌ریزی مود توجه قرار گرفت.

از سوی دیگر، بافتارهای استنادی که با انگیزه‌های استنادی «تصحیح اثر دیگران»، «رد اندیشه منتسب به دیگران» و «رد ادعاهای حق تقدم در آثار دیگران» ایجاد شده اند، شامل خوداستنادی نمی‌باشند؛ زیرا نویسنده به دیگران استناد داده است تا به آن وسیله، اثر دیگری را تصحیح کند یا اندیشه و ادعای او را رد نماید. لذا، خوداستنادی برای این گونه بافتارها لحاظ نشد.

جدول ۳. طرح حاشیه‌نویسی فراگفتمان‌مدار بافتارهای استنادی

نام برچسب	توضیحات برچسب
BKG-HDG	اطلاعات زمینه‌ای - احتیاطی
BKG-BOS	اطلاعات زمینه‌ای - تأکیدی
BKG-ATT	اطلاعات زمینه‌ای - نگرشی
BKG-SLF	اطلاعات زمینه‌ای - خوداستنادی
BKG-ENG	اطلاعات زمینه‌ای - مشارکت‌جویانه
LIT-REW-HDG	مرور پیشینه - احتیاطی
LIT-REW-BOS	مرور پیشینه - تأکیدی
LIT-REW-ATT	مرور پیشینه - نگرشی
LIT-REW-SLF	مرور پیشینه - خوداستنادی
LIT-REW-ENG	مرور پیشینه - مشارکت‌جویانه
PBM-ATT	بیان مسئله نگرشی
PBM-BOS	بیان مسئله تأکیدی
PBM-SLF	بیان مسئله با خوداستنادی
PBM-HDG	بیان مسئله احتیاطی
PBM-ENG	بیان مسئله مشارکت‌جویانه

نام برچسب	توضیحات برچسب
CONTI-HDG	ادامه و گسترش نتایج پژوهش قبلی نویسنده (خوداستنادی) - احتیاطی
CONTI-BOS	ادامه و گسترش نتایج پژوهش قبلی نویسنده (خوداستنادی) - تأکیدی
CONTI-ATT	ادامه و گسترش نتایج پژوهش قبلی نویسنده (خوداستنادی) - نگرشی
CONTI-ENG	ادامه و گسترش نتایج پژوهش قبلی نویسنده (خوداستنادی) - مشارکت‌جویانه
EXTEN-HDG	توسعه داده‌ها، روش‌ها و رویکردهای دیگر نویسندگان - احتیاطی
EXTEN-BOS	توسعه داده‌ها، روش‌ها و رویکردهای دیگر نویسندگان - تأکیدی
EXTEN-ATT	توسعه داده‌ها، روش‌ها و رویکردهای دیگر نویسندگان - نگرشی
EXTEN-ENG	توسعه داده‌ها، روش‌ها و رویکردهای دیگر نویسندگان - مشارکت‌جویانه
FUT-BOS	بیان پژوهش‌های آینده - تأکیدی
FUT-HDG	بیان پژوهش‌های آینده - احتیاطی
FUT-SLF	بیان پژوهش‌های آینده - خوداستنادی
FUT-ATT	بیان پژوهش‌های آینده - نگرشی
FUT-ENG	بیان پژوهش‌های آینده - مشارکت‌جویانه
RSL-ANL-ENG	تحلیل یافته‌ها - مشارکت‌جویانه
RSL-ANL-ATT	تحلیل یافته‌ها - نگرشی
RSL-ANL-BOS	تحلیل یافته‌ها - تأکیدی
RSL-ANL-HDG	تحلیل یافته‌ها - احتیاطی
RSL-ANL-SLF	تحلیل یافته‌ها - خوداستنادی
CON-COMP-SUP-ENG	مقایسه نتایج پژوهشی - ارائه شواهد همسوی مشارکت‌جویانه
CON-COMP-SUP-BOS	مقایسه نتایج پژوهشی - ارائه شواهد همسوی تأکیدی
CON-COMP-SUP-HDG	مقایسه نتایج پژوهشی - ارائه شواهد همسوی احتیاطی
CON-COMP-SUP-SLF	مقایسه نتایج پژوهشی - ارائه شواهد همسو با خود استنادی
CON-COMP-SUP-ATT	مقایسه نتایج پژوهشی - ارائه شواهد همسوی نگرشی
CON-COMP- DFF -ENG	مقایسه نتایج پژوهشی - ارائه شواهد غیرهمسوی مشارکت‌جویانه
CON-COMP- DFF -BOS	مقایسه نتایج پژوهشی - ارائه شواهد غیرهمسوی تأکیدی
CON-COMP- DFF -HDG	مقایسه نتایج پژوهشی - ارائه شواهد غیرهمسوی احتیاطی
CON-COMP- DFF -ATT	مقایسه نتایج پژوهشی - ارائه شواهد غیرهمسوی نگرشی
CON-COMP-DFF-SLF	مقایسه نتایج پژوهشی - ارائه شواهد غیرهمسو با خود استنادی
MTH-INT-HDG	معرفی روش و ابزارهای علمی مورد استفاده - احتیاطی
MTH-INT-BOS	معرفی روش و ابزارهای علمی مورد استفاده - تأکیدی
MTH-INT-ATT	معرفی روش و ابزارهای علمی مورد استفاده - نگرشی
MTH-INT-ENG	معرفی روش و ابزارهای علمی مورد استفاده - مشارکت‌جویانه
MTH-INT-SLF	معرفی روش و ابزارهای علمی مورد استفاده - خود استنادی
OWN-CRR-ENG	تصحیح اثر خود - مشارکت‌جویانه
OWN-CRR-BOS	تصحیح اثر خود - تأکیدی

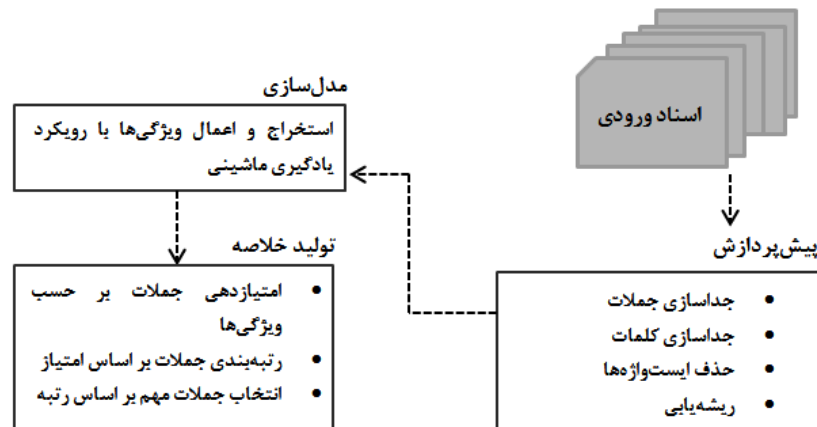
نام برچسب	توضیحات برچسب
OWN-CRR-HDG	تصحیح اثر خود-احتیاطی
OWN-CRR-ATT	تصحیح اثر خود-نگرشی
OWN-CRR-SLF	تصحیح اثر خود-خود استنادی
OTHER-CRR-ENG	تصحیح آثار دیگران - مشارکت‌جویانه
OTHER-CRR-BOS	تصحیح آثار دیگران - تأکیدی
OTHER-CRR-HDG	تصحیح آثار دیگران - احتیاطی
OTHER-CRR-ATT	تصحیح آثار دیگران - نگرشی
PROV-CLM-ENG	اثبات یک ادعا - مشارکت‌جویانه
PROV-CLM-BOS	اثبات یک ادعا - تأکیدی
PROV-CLM-HDG	اثبات یک ادعا - احتیاطی
PROV-CLM-ATT	اثبات یک ادعا - نگرشی
PROV-CLM-SLF	اثبات یک ادعا با خود استنادی
OTHER-REJ-ENG	رد اثر یا اندیشه منتسب به دیگران - مشارکت‌جویانه
OTHER-REJ-HDG	رد اثر یا اندیشه منتسب به دیگران - احتیاطی
OTHER-REJ-BOS	رد اثر یا اندیشه منتسب به دیگران - تأکیدی
OTHER-REJ-ATT	رد اثر یا اندیشه منتسب به دیگران - نگرشی
PRIOR-REJ-ENG	رد ادعاهای حق تقدم در آثار دیگران - مشارکت‌جویانه
PRIOR-REJ-HDG	رد ادعاهای حق تقدم در آثار دیگران - احتیاطی
PRIOR-REJ-BOS	رد ادعاهای حق تقدم در آثار دیگران - تأکیدی
PRIOR-REJ-ATT	رد ادعاهای حق تقدم در آثار دیگران - نگرشی

کاربرد طرح حاشیه‌نویسی پیشنهاد شده در خلاصه‌سازی خودکار استنادمدار

به طور کلی، خلاصه‌های خودکار به دو روش استخراجی^۱ و چکیده‌سازی^۲ تولید می‌شود. خلاصه‌های چکیده‌ای، تفسیری از متن اصلی را ارائه می‌کنند و در تولید آن‌ها مفاهیم جملات اصلی به شکل کوتاه‌تر بازنویسی می‌شوند (مانی^۳، ۲۰۰۱). خلاصه‌های استخراجی شامل مجموعه‌ای از جملات سند ورودی هستند. این نوع خلاصه‌سازی یک روش ساده اما قوی برای خلاصه‌سازی متن است. لذا بسیاری از پژوهش‌های این حوزه، از نوع استخراجی هستند. استخراج و انتخاب جملات مناسب برای خلاصه نهایی، بر اساس ویژگی‌های متون صورت می‌پذیرند. این ویژگی‌ها^۴، طیف وسیعی از ویژگی‌های سطحی مانند فراوانی کلمات و محل قرار

1. Extractive
2. Abstractive
3. Mani
4. Features

گرفتن جمله در متن تا ویژگی‌های معنایی را شامل می‌شوند (ساگیون و پویبو، ۲۰۱۳). این نوع از خلاصه‌سازی شامل ۳ مرحله اصلی است که عبارتند از پیش‌پردازش متن، مدل‌سازی و تولید خلاصه. نگاه کنید به شکل ۱.



شکل ۱. معماری کلی یک سامانه خلاصه‌ساز استخراجی

وجود پیکره مناسب در مرحله مدل‌سازی بسیار اهمیت دارد. زیرا، بر اساس آن می‌توان ویژگی‌های مفید برای تولید خلاصه‌های با کیفیت‌تر را استخراج کرد.

با اتخاذ رویکرد استخراجی در خلاصه‌سازی خودکار استنادمدار، لازم است به ویژگی‌های معنایی بافتارهای استنادی توجه شود. لذا، لازم است پیکره‌های مورد نیاز را تولید کرد.

یکی از ویژگی مهم در خلاصه‌سازی خودکار استنادمدار، کارکردهای استناد است. پیکره‌های موجود، کارکردهای استناددهی را بر اساس تحلیل معنایی بافتارها در سطح کلمه و جمله طبقه‌بندی کرده‌اند و در فرآیند خلاصه‌سازی خودکار تنها تلاش کرده‌اند با تعیین چهریزه‌گفتمانی بافتارهای استنادی مقالات استنادکننده خلاصه‌ای از مقاله استنادشونده را تولید کنند. برخی از این چهریزه‌ها عبارتند از: هدف، روش و نتیجه.

از آنجایی که در خلاصه‌سازی استنادمدار، تحلیل عمیق‌تر بافتارهای استنادی به منظور پالایش آن‌ها و تولید خلاصه‌های مفیدتر ضروری می‌نماید، بهره‌گیری از طرح پیشنهاد شده در این پژوهش می‌تواند برای ساخت پیکره مناسب راهگشا باشد. طرح پیشنهاد شده، منجر به استخراج ویژگی‌های فراگفتمانی بافتارهای استنادی می‌گردد.

در پیکره‌هایی که مبتنی بر طرح پیشنهاد شده، تولید شوند؛ به هر بافتار استنادی یک برچسب فراگفتمانی اختصاص داده می‌شود که بر اساس آن‌ها می‌توان الگوریتم‌های خلاصه‌سازی مبتنی بر رویکرد وزن‌دهی برچسب‌ها را ارائه داد. به این ترتیب، می‌توان برای بهره‌گیری از رویکرد غیردودویی وزن‌دهی نیز سیاست‌گذاری نمود. برای مثال، می‌توان قوانین وزن‌دهی را در الگوریتم تعریف کرد. برای نمونه، به برچسب‌هایی که تحلیل‌های نگرشی و تأکیدی نویسنده استناد کننده را آشکار می‌کنند، نسبت به برچسب‌هایی که به بافتارهایی با محتوای احتیاطی مربوط هستند، وزن بیشتری تخصیص داد. بدین ترتیب، خلاصه‌ساز قادر خواهد بود آن دسته از جملاتی که برچسب‌هایی با وزن بالاتر دارند را شناسایی نماید و در خلاصه نهایی درج کند.

علاوه بر کاربرد این طرح در امر خلاصه‌سازی خودکار استنادمدار، می‌توان از آن در پژوهش‌هایی که فقط بر خودکارسازی طبقه‌بندی کارکردهای استناد تمرکز دارند نیز بهره گرفت.

نتیجه‌گیری

تحلیل و پالایش بافتارهای استنادی از ملزومات خلاصه‌سازی خودکار استنادمدار است. زیرا، در این نوع از خلاصه‌سازی تلاش می‌شود خلاصه‌ای از بافتارهای استنادی مقالات استنادکننده به همراه خلاصه‌ای از مقاله استنادشونده به کاربر ارائه شود. بنابراین، اگر خلاصه‌ساز بتواند بافتارهای استنادی مفید را از استنادهای سراسری تشخیص دهد، می‌توان سهم نویسنده استنادشونده در علم را به صورت کیفی معلوم نمود و سپس برای درج آن‌ها در خلاصه استنادمدار نهایی تصمیم‌گیری نمود.

از این روی، رویکردهای معنایی خلاصه‌سازی خودکار استنادمدار عموماً از پیکره‌های حاشیه‌نویسی شده توسط متخصصان بهره می‌برند. ساخت چنین پیکره‌هایی به صرف هزینه کافی نیاز دارد و زمان‌بر است. فرآیند کار نیز، به این صورت است که یک طرح حاشیه‌نویسی در اختیار حداقل دو متخصص در حوزه مورد نظر، قرار می‌گیرد که برچسب‌زنی جملات را بر اساس آن طرح انجام می‌دهند. پیکره نهایی بر اساس توافق مطلوب حاشیه‌نویسان تولید می‌گردد.

در این مقاله به تحلیل نظری کاربرد الگوی فراگفتمان هایلند در خلاصه‌سازی خودکار استنادمدار متون علمی پرداخته شد و طرح حاشیه‌نویسی فراگفتمان مدار بافتارهای استنادی، پیشنهاد گردید. بررسی و تحلیل نظری متون نشان داد که از فراگفتمان تعاملی - مشارکتی هایلند می‌توان برای تحلیل بافتارهای استنادی بهره گرفت. لذا، طرح حاشیه‌نویسی بر اساس تردیدنما، یقین‌نما، نگرش‌نما، خود اظهار و دخیل‌سازها و با ۷۰ طبقه پیشنهاد شد.

با توجه به این که طرح حاشیه‌نویسی پیشنهاد شده مبتنی بر نظریات موجود است، لازم است در به کارگیری آن، از حاشیه‌نویسان خواسته شود تا در حین برچسب‌زنی، هر برچسب دیگری غیر از موارد مطرح شده در طرح را که به ذهنشان می‌رسد با ذکر دلیل، یادداشت نمایند تا در صورت احراز توافق مطلوب به طرح اضافه گردد.

پژوهشگران درصدد هستند، از طرح پیشنهاد شده برای تحلیل و پالایش بافتارهای استنادی مقالات حوزه علم اطلاعات و دانش‌شناسی استفاده کنند. همچنین، پیاده‌سازی یک سامانه خلاصه‌ساز استنادمدار برای مقالات علم اطلاعات در دست اقدام است که از پیکره حاشیه‌نویسی شده فراگفتمان‌مدار بافتارهای استنادی، بهره خواهد برد.

منابع

- تاجر، پگاه؛ جوکار، عبدالرسول؛ فخراحمد، سید مصطفی؛ خرمايي، علیرضا؛ ستوده، هاجر (۱۳۹۶). کاربرد تحلیل گفتمان در خلاصه‌سازی خودکار متون علمی. مجموعه مقالات نخستین همایش ملی رویکردهای نوین در مطالعات زبان و ادبیات، ۲۶ بهمن، مؤسسه آموزش عالی زند، شیراز، ایران.
- خانی، رضا؛ چنگیزی، مهسا (۱۳۹۴). مقایسه نشانه‌های فراگفتمان در مجلات انگلیسی بین‌المللی و داخلی: مطالعه موردی مقالات حوزه زبان‌شناسی کاربردی. *زبان‌پژوهی* (آماده انتشار).
- طارمی، طاهره؛ تاکی، گیتی؛ یوسفیان، پاکزاد (۱۳۹۷). جنسیت در مقالات علمی فارسی زبان: مطالعه پیکره بنیاد نشانگرهای فراگفتمان تعاملی بر اساس انگاره هایلند. *پژوهش‌های زبان‌شناسی*، ۱۰(۱)، ۲۳-۴۱.
- قدیمی، آفتاب؛ ستوده، هاجر (۱۳۹۵). بررسی جایگاه استنادی مقالات شیمی کشور ایران در نمایه استنادی علوم. *پژوهش‌نامه علم‌سنجی*، ۲(۴)، ۶۳-۷۴.
- قلی‌فامیان، علی‌رضا؛ کارگر، مریم (۱۳۹۳). تحلیل مقالات نقد کتاب‌های زبان‌شناسی ایران بر اساس الگوی فراگفتمان هایلند. *پژوهش‌های زبان‌شناسی*، ۵(۲)، ۳۷-۵۲.

References

- Abu-Jbara, A., & Radev, D. (2011, June). Coherent citation-based summarization of scientific papers. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1* (pp. 500-509). Association for Computational Linguistics.
- Athar, A. (2014). *Sentiment analysis of scientific citations*. University of Cambridge, Computer Laboratory.
- Baxendale, P. B. (1958). Machine-made index for technical literature - an experiment. *IBM Journal of Research and Development*, 2(4), 354-361.

- Cohan, A., Soldaini, L., & Goharian, N. (2014). Towards Citation-based summarization of biomedical literature. In *Proceedings of the Text Analysis Conference (TAC'14)*.
- Elkiss, A., Shen, S., Fader, A., Erkan, G., States, D., & Radev, D. (2008). Blind men and elephants: What do citation summaries tell us about a research article? *Journal of the American Society for Information Science and Technology*, 59(1), 51-62.
- Fisas, B., Ronzano, F., & Saggion, H. (2015, June). On the discursive structure of computer graphics research papers. In *The 9th Linguistic Annotation Workshop held in conjunction with NAACL 2015* (p. 42).
- Ghadimi, A., & Sotoudeh, H. (2016). Investigating the Citation Location of Iranian Chemistry articles in Science Citation Index. *Scientometrics Research*, 2(4), 63-74. (in Persian).
- Garfield, E. (1965, December). Can citation indexing be automated. In *Statistical association methods for mechanized documentation, symposium proceedings*. 269, 189-192. Washington, DC: National Bureau of Standards, Miscellaneous Publication 269.
- Gholifamian, A., & Kargar, M. (2014). An analysis of review articles of Iranian linguistic books based on Hyland metadiscourse model. *Journal of Researches in Linguistics*, 5(2), 37-52. (in Persian)
- Hernández-Alvarez, M., & Gómez, J. M. (2016). Survey about citation context analysis: Tasks, techniques, and resources. *Natural Language Engineering*, 22(3), 327-349.
- Hyland, K. (1998). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of pragmatics*, 30(4), 437-455.
- Hyland, K. (2005a). *Metadiscourse: Exploring Interaction in Writing*. London: Continuum.
- Hyland, K. (2005b). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192.
- Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied linguistics*, 25(2), 156-177.
- Hyland, K. (2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of second language writing*, 13(2), 133-151.
- Khani, R., & Changizi, M. (2015). Comparison of metadiscourse markers between international and national English journals: A case of applied linguistics articles. *Journal of Language Research*, (Ready to Publish). (in Persian)
- Kupiec, J., Pedersen, J., & Chen, F. (1995, July). A trainable document summarizer. In *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 68-73). ACM.
- Jurgens, D., Kumar, S., Hoover, R., McFarland, D., & Jurafsky, D. (2016). Citation Classification for behavioral analysis of a scientific field. *arXiv preprint arXiv:1609.00435*
- Liakata, M., Teufel, S., Siddharthan, A., & Batchelor, C. R. (2010, May). Corpora for the Conceptualisation and Zoning of Scientific Papers. In *LREC*.
- Liakata, M., Saha, S., Dobnik, S., Batchelor, C., & Rebolz-Schuhmann, D. (2012). Automatic recognition of conceptualization zones in scientific articles and two life science applications. *Bioinformatics*, 28(7), 991-1000.

- Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2), 159-165.
- Mani, I. (2001). *Automatic summarization* (Vol. 3). John Benjamins Publishing.
- Qazvinian, V., & Radev, D. R. (2008, August). Scientific paper summarization using citation summary networks. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1* (pp. 689-696). Association for Computational Linguistics.
- Qazvinian, V., & Radev, D. R. (2010, July). Identifying non-explicit citing sentences for citation-based summarization. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 555-564). Association for Computational Linguistics.
- Qazvinian, V., Radev, D. R., Mohammad, S. M., Dorr, B., Zajic, D., Whidby, M., & Moon, T. (2013). Generating extractive summaries of scientific paradigms. *Journal of Artificial Intelligence Research*. 46, pp.165-201.
- Saggion, H., and Poibeau, T. (2013). *Automatic text summarization: Past, present and future*. In *Multi-source, multilingual information extraction and summarization* (pp. 3-21). Springer Berlin Heidelberg.
- Tajer, P., Jowkar, A., Fakhrahmad, S. M., Khormae, A., & Sotoudeh, H. (2018). Application of Discourse Analysis in Automatic Summarization of Scientific Texts. *Proceedings of the First National Conference on New Approaches to Language and Literature Studies*, February 15, Zand Institute of Higher Education, Shiraz, Iran. (in Persian).
- Taremi, T., Taki, G., & Yousefian, P. (2018). Gender in Farsi scientific articles: a corpus-based study of interactive metadiscourse markers based on Hyland model. *Journal of Researches in Linguistics*, 10(1). 23-41. (in Persian).
- Teufel, S. (1999). *Argumentative Zoning: Information Extraction from Scientific Text*. School of Cognitive Science, University of Edinburg, UK.
- Teufel, S., & Moens, M. (2002). Summarizing scientific articles: experiments with relevance and rhetorical status. *Computational linguistics*, 28(4), 409-445.
- Teufel, S. (2010). *The structure of scientific articles: Applications to citation indexing and summarization*: Center for the Study of Language and Information.
- Teufel, S., Siddharthan, A., & Batchelor, C. (2009, August). Towards discipline-independent argumentative zoning: evidence from chemistry and computational linguistics. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*. 3, 1493-1502. Association for Computational Linguistics.
- Teufel, S., Siddharthan, A., & Tidhar, D. (2006, July). Automatic classification of citation function. In *Proceedings of the 2006 conference on empirical methods in natural language processing*, 103-110. Association for Computational Linguistics.
- Yeloglu, O., Milios, E., & Zincir-Heywood, N. (2011, March). Multi-document summarization of scientific corpora. In *Proceedings of the 2011 ACM Symposium on Applied Computing*, 252-258. ACM.