



ساخت پیکره متنی: طراحی مدل امکان‌سنجی

الهام علایی ابوذری^{*۱}

سیروس علیدوستی^۲

مقاله پژوهشی

چکیده

ساخت پیکره‌ها کاری پیچیده، زمان‌بر، دارای گام‌های گوناگون، و میان‌رشته‌ای است. بنابراین برای کارایی و اثربخشی در ساخت آنها نیاز است که پیش از آغاز کار، امکان‌سنجی انجام شود. امکان‌سنجی؛ گام‌ها، هزینه‌ها، نیروی انسانی، حقوق مادی و معنوی، و مانند آنها را برای یک پروژه، بررسی و به مدیریت پروژه کمک می‌کند تا آن را با آمادگی و آینده‌نگری بیشتری به سرانجام برساند. از آنجایی که مدلی برای امکان‌سنجی ساخت پیکره متنی همچون گونه‌ای از پیکره‌ها در پژوهش‌های پیشین نبود، این پژوهش در پی ساخت چنین مدلی انجام شد. برای این کار، نخست پژوهش‌های پیشین ساخت پیکره متنی و امکان‌سنجی بررسی و بر این پایه یک مدل پیشنهاد شد. سپس با روش دلفی و در دو دور اعتباریابی انجام شد و بر این پایه مدل پایانی به همراه ابزارها و حاشیه‌نویسی‌ها برای آن به دست آمد. این مدل دارای هفت بُعد فنی، اقتصادی، زمان‌بندی، قانونی حقوقی، عملیاتی، تأمین مالی و بازاریابی و روی‌هم‌رفته ۳۳ مؤلفه است.

کلیدواژه‌ها: پیکره زبانی، پیکره متنی، امکان‌سنجی، مدل، روش دلفی.

✉ Alayi@irandoc.ac.ir

✉ alidousti@irandoc.ac.ir

۱- استادیار پژوهشگاه علوم و فناوری اطلاعات ایران (ایراندک)*

۲- دانشیار پژوهشگاه علوم و فناوری اطلاعات ایران (ایراندک)

۱- مقدمه

در سال‌های گذشته پژوهشگران بسیاری به ساخت پیکره‌های^۱ زبانی و واکاوی پیکره‌بنیاد زبان پرداخته‌اند. پیکره مجموعه‌ای نظام‌مند، رایانه‌ای، معتبر^۲ و موثق^۳ از زبان است که در پژوهش‌های زبان‌شناسی به کار می‌رود. مجموعه داده‌های زبانی در پیکره می‌تواند متن و هم برگردان نوشتاری از گفتار ضبط شده باشد. هر مجموعه‌ای را که بیش از یک متن داشته باشد، می‌توان پیکره متنی^۴ خواند. «پیکره» یا «پیکره متنی» بیشتر در زبان‌شناسی نوین به کار می‌روند و به مجموعه‌ای از متون گفته می‌شوند که برای واکاوی زبان‌شناسی یا پردازش زبان طبیعی به کار می‌روند. مک‌انری و ویلسون (۲۰۰۱) پیکره را مجموعه‌ای کمابیش بزرگ از متون الکترونیکی می‌خوانند که حاشیه‌نویسی، برچسب‌گذاری و دسته‌بندی سنجیده‌ای دارند و از همین رو نیز بررسی‌های بسیار گوناگونی را برای کاربر شدنی می‌سازند. از دهه ۱۹۶۰ به این سو، پژوهشگران توانستند با کمک رایانه، مجموعه‌های بسیار بزرگی از متون معتبر زبانی را فراهم کنند که آغازگر همه این‌ها، پیکره زبانی «براون»^۵ بود. این پیکره، نمونه‌ای معیار از زبان انگلیسی امروز برای کاربرد در رایانه‌های رقمی به‌شمار می‌رود (اسوارت ویکل، ۱۹۹۲، نقل از عاصی، ۱۳۸۲). در آغاز دهه ۱۹۸۰ شاید می‌شد پیکره‌های الکترونیکی را با انگلستان یک دست شمرد. ولی در آغاز دهه ۱۹۹۰ شمار طرح‌هایی که در پی ساخت پیکره‌های زبانی بودند به ده‌ها می‌رسید. اکنون پس از گذشت تنها چند دهه، با جست‌وجویی در اینترنت بیش از سیصد پیکره بزرگ و کوچک برای بسیاری از زبان‌های جهان یافت می‌شوند (عاصی، ۱۳۸۲). در ایران نیز دانشگاه‌ها و پژوهشگاه‌های گوناگون ساخت پیکره را آغاز کرده‌اند که از میان آنها می‌توان از پیکره «همشهری» (آل احمد و همکاران، ۲۰۰۹)، پیکره «فارس‌نت» (شمس‌فرد و همکاران، ۲۰۱۰)، پیکره متنی زبان فارسی که توسعه یافته پیکره «بی‌جن‌خان» است (بی‌جن‌خان و همکاران، ۲۰۱۱)؛ پیکره وابستگی نحوی زبان فارسی (رسولی و همکاران، ۲۰۱۳)، پیکره چندزبانه رایانامه‌ها (دهقانی و همکاران، ۲۰۱۳)، پیکره واژگان فارسی دارای برچسب قطبیت (دهداربهبهانی و همکاران، ۲۰۱۴)، پایگاه دادگان زبان فارسی (عاصی، ۱۳۸۴)، پیکره نقش‌های معنایی زبان فارسی (میرزایی و مولودی، ۱۳۹۳) و پیکره تشخیص خودکار جنسیت (مرادی و بحرانی، ۱۳۹۴) نام برد. ویژگی‌های برخی از این پیکره‌ها در جدول یک آمده است:

1. corpus
2. authentic
3. reliable
4. text corpus
5. Brown Corpus

جدول ۱: مشخصات برخی از بیکره‌های فارسی

پیکره	ویژگی‌ها	زمان
پیکره فارسی‌نت	فارسی‌نت در آزمایشگاه پردازش زبان طبیعی دانشگاه شهیدبهشتی با پشتیبانی مرکز تحقیقات مخابرات ایران ساخته شده است. واپسین ویراست فارسی‌نت (نسخه ۳) دارای بیش از ۱۰۰ هزار مدخل واژگانی (واژه یا عبارت) است که در ۴۰ هزار مجموعه ترادف جا گرفته‌اند. هر مدخل نیز دست کم یک معنی دارد که در یک و تنها یک مجموعه ترادف شرکت می‌کند. همه مجموعه‌های ترادف یا در سلسله مراتب شرکت می‌کنند یا سرگروه هستند. هر مجموعه ترادف نیز یا دست کم یکی از اعضا آن در دست کم یک رابطه غیرسلسله مراتبی شرکت کرده است. هر مجموعه ترادف به مجموعه ترادف همانند در وردنت ۳٫۰ پرینستون نگاشت شده است (farsnet.nlp.sbu.ac.ir).	دادگان فارسی‌نت در چهار پروژه تهیه شده‌اند. فارسی‌نت ۱ در سال ۱۳۸۷ ساخته شد. فارسی‌نت ۲ در سال ۱۳۸۹ و در سال ۱۳۹۶ هم پروژه فارسی‌نت ۳ با هدف رسیدن به ۱۰۰۰۰۰ مدخل در ۴۰۰۰۰ هزار مجموعه ترادف به انجام رسید. نسخه اول فارسی‌نت شامل بیش از ۱۷ هزار مدخل واژگانی از مقوله‌های اسم، فعل و صفت است. روابط تحت پوشش در این نسخه روابط درون‌مقوله‌ای مطرح در وردنت انگلیسی (نسخه ۲/۱) می‌باشد و قابلیت اتصال به وردنت‌های دیگر از طریق نگاشت به وردنت پرینستون نسخه ۳/۰ را نیز داراست. نسخه دوم فارسی‌نت شامل بیش از ۳۰ هزار مدخل واژگانی از مقوله‌های اسم، فعل، صفت و قید است. علاوه بر روابط درون‌مقوله‌ای مطرح در وردنت انگلیسی (نسخه ۲/۱)، پنج رابطه میان مقوله‌های نیز م‌فاهیم را بهم پیوند می‌دهد و علاوه بر ویژگی‌های در نظر گرفته شده برای واژه‌ها، ویژگی‌های نحوی، ساخت‌واژی و آوایی به واژه‌ها و قاب و ساختار آرگومانی به افعال افزوده شده است. این وردنت نیز قابلیت اتصال به وردنت‌های دیگر را از طریق نگاشت به وردنت پرینستون نسخه ۳/۰ داراست. نسخه دوم و نیم فارسی‌نت شامل بیش از ۳۳ هزار مدخل واژگانی از مقوله‌های اسم، فعل، صفت و قید است. این نسخه نگارش بهبود یافته و تکمیل شده نسخه دوم می‌باشد. نسخه سوم فارسی‌نت شامل بیش از ۱۰۰ هزار مدخل واژگانی در ۴۰۰۰۰ هزار مجموعه ترادف می‌باشد. با توجه به وردنت‌های بزرگی که در دنیا وجود دارد و تجربه پژوهش‌های انجام‌شده گسترش فارسی‌نت از جهت دامنه پوشش لغات فارسی و هم تعداد روابط بسیار اهمیت دارد. در پروژه توسعه فارسی‌نت ۳ هدف توسعه و تکمیل هستان‌شناسی عمومی زبان فارسی تا ۱۰۰ هزار مدخل واژگانی بود. همچنین با بازبینی دادگان قبلی تضمین دقت ۹۰ درصد روی دادگان فارسی‌نت ۲ از دیگر اهداف پروژه بود. (farsnet.nlp.sbu.ac.ir).

<p>پیشینه‌ای یافت نشد.</p>	<p>پیکره متنی زبان فارسی، مجموعه‌ای از متون نوشتاری و گفتاری رسمی زبان فارسی است که از منابع واقعی همچون روزنامه‌ها، سایت‌ها و مستندات از پیش تایپ شده؛ گردآوری و ویرایش شده و برچسب خورده است. این دادگان نزدیک به ۱۰۰ میلیون واژه دارد که از منابع گوناگون فراهم شده و دارای گوناگونی بسیاری است. ۱۰ میلیون کلمه از این پیکره با ۸۸۲ برچسب نحوی معنایی به دست دانشجویان رشته زبان‌شناسی برچسب‌دهی و هر پرونده بر حسب موضوع و منبع آن طبقه‌بندی شده است. این پیکره را پژوهشکده پردازش هوشمند علائم ساخته و برای کاربرد در آموزش مدل زبانی و پروژه‌های پردازش زبان طبیعی مناسب است (بی‌جن‌خان و همکاران، ۲۰۱۱).</p>	<p>پیکره متنی زبان فارسی/ نسخه کامل پیکره بی‌جن‌خان</p>
<p>پیشینه‌ای یافت نشد.</p>	<p>پیکره وابستگی نحوی زبان فارسی، مجموعه‌ای با ۳۰ هزار جمله برچسب‌خورده با اطلاعات نحوی و ساخت‌واژی است. جمله‌های پیکره برگرفته از منابع گوناگون نوشتارهای فارسی معاصر هستند. همه جمله‌های پیکره دارای برچسب روابط نحوی (بر پایه دستور وابستگی) مانند فاعل، مفعول، مسند، مضاف‌الیه، بدل و همچنین برچسب اطلاعات ساخت‌واژی (برچسب اجزای واژگانی کلام (POS)) مانند فعل، اسم، صفت، قید، ضمیر، و... هستند. جمله‌ها را گروهی از زبان‌شناسان مجرب برچسب‌زده و چند بار بازبینی کرده‌اند. داده‌های پیکره، تصادفی به داده‌های یادگیری (۸۰٪)، آزمون (۱۰٪) و ارزیابی (۱۰٪) تقسیم شده‌اند. شماره همه جمله‌های پیکره ۲۹۹۸۲، شماره همه واژه‌ها ۴۹۸۰۸۱، شمار واژه‌های یکتا ۳۷۶۱۸، میانگین هر جمله ۱۶/۶۱ شمار فعل‌های یکتا ۴۷۸۲ و میانگین دیده شدن هر فعل ۱۲/۶۷ است (رسولی و همکاران، ۲۰۱۳).</p>	<p>پیکره وابستگی نحوی زبان فارسی</p>

پایگاه دادگان زبان فارسی	<p>پایگاه داده‌های زبان فارسی مجموعه‌ای است از متون گوناگون فارسی که بخشی از آن دارای نشانه‌گذاری‌هایی مانند شناسنامه متن، برچسب‌های دستوری، آوایی، ریشه‌ای و معنایی است. این دادگان که در پژوهشگاه علوم انسانی و مطالعات فرهنگی درست شده و مجهز به نرم‌افزارهای ویژه جست‌وجو، تقطیع و تحلیل متن است که می‌تواند انواع فهرست‌های واژگانی، بسامدی و آماری را ارائه کند. از جمله متون موجود در پیکره، نزدیک ۴۵۰ اثر داستانی و غیرداستانی نثر؛ ۲۵۰ اثر شعری از شاعران معاصر؛ بیش از ۸۰ عنوان مجله و نشریه علمی، ادبی، و تخصصی؛ نزدیک به ۳۰۰ عنوان نمایشنامه و فیلمنامه؛ ۲۰۰ عنوان ادبیات کودک؛ چندین عنوان روزنامه و نشریه خبری؛ برخی از کتاب‌های درسی دانشگاهی و دبیرستانی؛ برخی از کتاب‌های دبستانی؛ نامه‌های اداری و بخشنامه‌ها؛ مجموعه کامل قوانین و مقررات؛ نشریه‌ها و جزوه‌های پراکنده؛ پوسترها و دیوارنوشته‌ها است (peykaregan.ir).</p>	<p>از اوایل سال ۱۳۷۲ کار ایجاد پایگاه داده‌ای برای زبان فارسی با طراحی و سرپرستی دکتر مصطفی عاصی در پژوهشگاه علوم انسانی آغاز شد و تا سال ۱۳۷۸ دو مرحله (فاز) آن به اجرا درآمد و مرحله سوم که مهم‌ترین فاز یعنی گسترش و افزایش حجم داده‌ها و دگرگونی اساسی در نرم‌افزار و ایجاد امکانات نوین شبکه‌ای برای ارائه خدمات و اطلاعات آن در شبکه جهانی اینترنت بود، بدلیل نبود منابع مالی چندسالی از اجرا بازماند تا اینکه با کمک مالی وزارت ارتباطات و فناوری اطلاعات از سال ۱۳۸۱ اجرای فاز سوم طرح آغاز گردید و دو سال بعد به پایان رسید (عاصی، ۱۳۸۴).</p>
-----------------------------------	---	--

زبان‌شناسی پیکره‌ای به مطالعه فقره‌های واژگانی، ساخت‌های دستوری، یا پیوند این دو با دیگر مشخصه‌های زبانی و غیرزبانی می‌پردازد. زبان‌شناسی پیکره‌ای را می‌توان گذار از رویکردهای ذهن‌گرایانه به پژوهش‌های زبانی دانست. پژوهش‌های پیکره‌ای بر الگوهای واقعی کاربرد زبان در پایگاه‌های بزرگ دادگان یا همان پیکره‌ها استوار هستند. در این میان، هم از روش‌های کمی و هم از روش‌های کیفی برای تحلیل زبان بهره می‌برد و رایانه را برای انجام تحلیل‌های پیچیده به کار می‌گیرد. آنچه زبان‌شناسی پیکره‌ای را به کار زبان می‌آورد، توانمندی آن در بررسی کاربرد واقعی زبان است تا پژوهشگر از زیاده‌روی در نظریه‌پردازی دور نگه داشته شود. همین توانمندی نیز واکاوی پایگاه‌های بزرگ داده‌های زبانی را به زبان‌شناسی پیکره‌ای می‌سپارد (کوهستانی، ۱۳۸۹). توبرت (۲۰۰۵) برای روشن شدن زبان‌شناسی پیکره‌ای در برابر سنتی، ویژگی‌هایی از شیوه مطالعات زبانی در چارچوب زبان‌شناسی پیکره‌ای را بر شمرده است که تأکید روی معنا، تأکید روی متن، کاربرد داده‌های واقعی زبان و توجه به بُعد در زمانی کلام را در بر دارد. زبان‌شناسی پیکره‌ای، ابزارهایی بنیادی برای بررسی‌ها، پژوهش‌ها و کاربردهای زبانی گوناگون فراهم می‌آورد. برخی از کاربردهای پیکره عبارتند از:

۱. پردازش زبان طبیعی^۱ و درک و بازشناسی گفتار (عاصی، ۱۳۸۲):
 ۲. ابزاری کارا و ناگزیر در تدوین فرهنگ‌ها (هارتمن، ۱۹۹۴، نقل از عاصی، ۱۳۸۲):
 ۳. منبعی کلیدی برای آموزش و پژوهش (السولیتی و اتول، ۲۰۰۶):
 ۴. ساخت پایگاه‌های داده زبانی (عاصی، ۱۳۸۲):
 ۵. بررسی واژه‌های همایند^۲ در زبان‌های گوناگون (عاصی، ۱۳۸۲):
 ۶. پایشگری زبان برای پیگیری و ردگیری دگرگونی‌های زبانی (کندی، ۱۹۹۸، نقل از عاصی، ۱۳۸۲):
 ۷. ترجمه ماشینی (عاصی، ۱۳۸۲):
 ۸. توسعه مفاهیم و منابع در پیوند با واژگان (لارنس، ۲۰۱۷):
 ۹. نگارش و گسترش مهارت‌های نوشتاری (فردریکوا کلیموا، ۲۰۱۴):
 ۱۰. آموزش و یادگیری زبان با شناخت گویش‌ها و گوناگونی زبانی، معناشناسی، تحلیل کلام، زبان‌شناسی اجتماعی، زبان‌شناسی حقوقی، واکاوی ژانرهای ادبی و پژوهش‌های دستوری (اسوارت‌ویکل، ۱۹۹۲، نقل از عاصی، ۱۳۸۲).
- پژوهشگرانی که به واکاوی پیکره زبانی می‌پردازند، شاید در ساخت پیکره یا پیکره‌هایی که به کار می‌برند، سببستگی نداشته باشند. از سویی نیز پژوهشگرانی که پیکره ساخته‌اند، چگونگی انجام آن را به خوبی ارائه نکرده‌اند. برخی از این پژوهشگران، تنها از چالش‌هایی سخن گفته‌اند که هنگام ساخت پیکره با آنها روبه‌رو شده‌اند. برخی از این چالش‌ها درباره زبان پیکره بوده است (مانند چالش‌های خط و زبان فارسی). ولی ساخت پیکره متنی پیچیدگی‌هایی (مانند گونه و اندازه متون، نمایندگی متون گزیده شده، توازن و...) دارد که باید در فرایند ساخت آن دیده شوند. ساخت پیکره‌ها کاری پیچیده، زمان‌بر، دارای گام‌های گوناگون و میان‌رشته‌ای است. بنابراین برای کارایی و اثربخشی در ساخت آنها نیاز است که پیش از آغاز کار، امکان‌سنجی انجام شود. امکان‌سنجی؛ گام‌ها، هزینه‌ها، نیروی انسانی و مانند آنها را برای یک پروژه، بررسی و به مدیریت پروژه کمک می‌کند تا آن را با آمادگی و آینده‌نگری بیشتری به سرانجام برساند. از آنجایی که مدلی فراگیر برای امکان‌سنجی ساخت پیکره متنی همچون گونه‌ای پرکاربرد از پیکره‌ها در پژوهش‌های پیشین نبود، این پژوهش در پی ساخت چنین مدلی انجام شد. این مدل باید مسئله کلیدی، و نیازها و راه‌حل‌های اولیه پروژه را نمایان کند و نشان دهد که آیا می‌توان با داشته‌ها از پس چالش‌ها برآمد و پروژه را به انجام رساند یا نه؟ (پاور و همکاران، ۱۹۹۰).
- بر این پایه، هدف این پژوهش ساخت مدلی فراگیر و معتبر برای ساخت پیکره متنی و شناخت ابزارها و حاشیه‌نویسی‌های گوناگون برای آن است. در بخش‌های دیگر مقاله، بررسی پژوهش‌های پیشین ساخت پیکره متنی و امکان‌سنجی و مدل به‌دست آمده، آورده شده است. سپس کاربرد روش دلفی برای اعتباریابی این مدل و مدل پایانی آمده است.

1. natural language processing

2. collocations

۲- بررسی پیشینه پژوهش

از آنجایی که بررسی پژوهش‌های پیشین نشان داد که کاری یکپارچه در زمینه امکان‌سنجی ساخت پیکره متنی انجام نشده است، بررسی پیشینه پژوهش در دو زمینه همسایه این موضوع، پژوهش‌های امکان‌سنجی و پژوهش‌های ساخت پیکره انجام شد که با آمیختن آنها بتوان زیربنای گام‌های دیگر پژوهش را فراهم ساخت.

۱-۲- پژوهش‌های امکان‌سنجی

امکان‌سنجی^۱ تجزیه و تحلیل قابلیت به عمل نشستن یک ایده در یک مسیر مشخص و منطقی و نظم و ترتیب دادن به مراحل انجام کار از ابتدا تا انتها است (توماس و همکاران، ۱۹۸۲، نقل از شگری، ۱۳۹۴). در امکان‌سنجی؛ هدف و نیازمندی‌های پروژه و راهکارهای آن بررسی و برای انجام آن توصیه‌هایی می‌شود. این توصیه‌ها نشان می‌دهند که آیا پروژه باید ادامه یابد یا نه؟ امکان‌سنجی هزینه‌های گوناگون انجام یک پروژه را برآورد می‌کند (پاور و همکاران، ۱۹۹۰). امکان‌سنجی یک روش مطالعه برای سنجش سودمندی، مزیت‌ها و هزینه‌های انجام یک پروژه است. هزینه و فایده، تنها معنای مادی و مالی ندارد، بلکه دارای نگاهی فراگیر و چند بُعدی (مالی، انسانی، سازمانی، قانونی، امکانات نرم‌افزاری و سخت‌افزاری) است که باید در کنار هم دیده شوند. امکان‌سنجی در پی شناخت وضع کنونی و امکان انجام طرح است. بنابراین تأثیر عملیاتی کردن و دگرگونی‌های پدید آمده از انجام یک ایده نوین را می‌سنجد. اگر ایده‌ای برای انجام مناسب دیده شود، امکان‌سنجی، اطلاعات تازه‌ای را درباره تأمین مالی، جزئیات هزینه و مانند آنها فراهم می‌کند (سالاری، ۱۳۹۵). در بررسی پژوهش‌های پیشین، پژوهشی درباره امکان‌سنجی ساخت پیکره زبانی یافت نشد. ولی از پژوهش‌هایی که در زمینه امکان‌سنجی انجام شده‌اند، می‌توان از پژوهش موخری و ری (۲۰۱۷) نام برد. آنها پنج زمینه/بُعد کلیدی امکان‌سنجی پروژه‌ها را پیشنهاد می‌کنند که عبارت‌اند از امکان‌سنجی فنی، امکان‌سنجی اقتصادی، امکان‌سنجی حقوقی، امکان‌سنجی عملیاتی و امکان‌سنجی زمان‌بندی. آنها در پایان نیز مؤلفه‌های یک امکان‌سنجی خوب و موفقیت‌آمیز را پیشنهاد می‌کنند. این مؤلفه‌ها عبارت‌اند از: اطمینان از درستی اطلاعاتی که در امکان‌سنجی به کار می‌رود، بیان درست و دقیق مسئله و اطلاعات مستند از وضعیت مالی برای انجام پروژه. در یک امکان‌سنجی خوب باید مخاطرات پیش و پس از انجام پروژه و احتمال بازگشت سرمایه را نیز بررسی کرد. بوس و همکاران (۲۰۱۴) نیز امکان‌سنجی را از پنج بُعد بررسی کرده‌اند: فنی، اقتصادی، قانونی حقوقی، عملیاتی و زمان‌بندی. آنها کوشیده‌اند پیوند میان امکان‌سنجی فنی و توسعه محصول را نشان دهند و برای همین به شناخت هدف‌های امکان‌سنجی فنی و نمونه‌هایی از آن در فرایند توسعه محصول توجه کرده‌اند. یانگ و همکاران (۲۰۱۲) از دستاوردهای امکان‌سنجی اقتصادی یک مزرعه بادی گفته‌اند که در آن سه عامل هزینه واکاوی وضع موجود، یارانه دولت برای مزرعه‌های بادی و سازوکار ساخت پاک^۲ مزرعه‌های بادی آمده است. جواهری (۱۳۸۳) با انجام امکان‌سنجی، کوشیده است که نگاهی واقع‌بینانه به جا،

1. feasibility study
2. clean development

مشتریان بالقوه، منابع مالی و توسعه فعالیت‌های تفریحی داشته باشد. اهمیت و ضرورت امکان‌سنجی، معرفی فعالیت‌های تفریحی و گردشگری، گام‌های امکان‌سنجی، شیوه‌ها و شاخص‌های در پیوند با این مطالعه در مقاله وی آمده است. او منابع و عوامل مکانی، سیمای مصرف‌کننده، سیمای ترکیبی مصرف‌کننده و فعالیت‌های تفریحی و گردشگری را با تخمین هزینه، درآمد و سود، ارزیابی و راهنمای امکان‌سنجی پروژه‌های تفریحی و گردشگری را هم پیشنهاد کرده است. ادیب‌منش و همکاران (۱۳۹۰) برنامه‌ریزی درسی مدرسه‌محور را در دبیرستان‌های شهر کرمانشاه امکان‌سنجی کرده‌اند. پژوهش آنها توصیفی است و از پرسشنامه برای گردآوری دیدگاه کارشناسان بهره برده‌اند. در این امکان‌سنجی چهار متغیر دانش، نگرش، مهارت و امکانات از دیدگاه دو گروه مدیران و دبیران سنجش و ارزیابی شده‌اند. می‌توان گفت که متغیرهای این امکان‌سنجی در دو بُعد فنی و عملیاتی بوده‌اند. پورفرج و همکاران (۱۳۹۱) سرمایه‌گذاری در شهرک تفریحی توریستی بین‌المللی طارم (در استان زنجان) را با سنجش‌های گوناگون، ارزیابی مالی و اقتصادی و با نرم‌افزار «کامفار»^۱ شاخص‌های اقتصادی آن را ارزیابی کرده‌اند. امکان‌سنجی آنها فنی، اقتصادی و بازاریابی است تا بتوان درباره سرمایه‌گذاری بر چنین پروژه‌ای تصمیم بهتری گرفت. در پژوهش نجفی‌خواه و همکاران (۱۳۹۵) امکان‌سنجی ارائه خدمات مرجع مجازی کتابخانه‌های عمومی استان گیلان وابسته به نهاد کتابخانه‌های عمومی کشور انجام شده است. در این کار، امکان‌سنجی فنی از چهار بُعد سخت‌افزار، نرم‌افزار، شبکه و منابع و پایگاه‌های اطلاعات الکترونیکی انجام شده است. در این پژوهش، وضعیت مالی برای انجام پروژه نیز از پنج بُعد هزینه نرم‌افزار و پشتیبانی، هزینه نیروی انسانی، هزینه آموزش، هزینه سخت‌افزار و هزینه منابع و پایگاه‌های اطلاعاتی الکترونیکی امکان‌سنجی شده است.

ابعاد و مؤلفه‌هایی را که در پژوهش‌های پیشین درباره امکان‌سنجی پیدا شده‌اند، می‌توان در جدول یک

دید.

جدول ۲: ابعاد و مؤلفه‌های امکان‌سنجی

ابعاد	معیارها/ مقیاس‌ها	منابع
۱. فنی	۱-۱. مهارت‌ها و تخصص‌های فنی و دانش لازم برای ساخت/ توسعه/ کاربست سیستم ۲-۱. فناوری و فنون لازم برای ساخت/ توسعه/ کاربست سیستم ۳-۱. کمک‌های بیرون از سازمان برای کاربرد سیستم تازه یا پشتیبانی از فناوری و فنون ۴-۱. طراحی زیرساخت پروژه ۵-۱. پیش‌نیازهای معماری و مهندسی پروژه ۶-۱. سخت‌افزار، نرم‌افزار، شبکه و منابع و پایگاه‌های اطلاعاتی الکترونیکی	(ویتنی، ۲۰۰۷، نقل از صرافی‌زاده و علیپور، ۱۳۹۰)، موخری و ری (۲۰۱۷)، بوس و همکاران (۲۰۱۴)، یانگ و همکاران (۲۰۱۲)، کرنقان (۲۰۱۲)، ادیب‌منش و همکاران (۱۳۹۰)، پورفرج و همکاران (۱۳۹۱)، نجفی‌خواه و همکاران (۱۳۹۵)،

ابعاد	معیارها/ مقیاس‌ها	منابع
۲. اقتصادی	<p>۱-۲. هزینه نرم‌افزار و سخت‌افزار</p> <p>۲-۲. هزینه مشاوره طرح/ پروژه</p> <p>۳-۲. هزینه‌های پنهان مانند زمان مجریان، مطالعه مقدماتی و آموزش</p> <p>۴-۲. هزینه واکاوی وضع موجود</p> <p>۵-۲. منابع مالی لازم برای پیاده‌سازی سیستم</p> <p>۶-۲. هزینه نیروی انسانی/ تیم اجرایی</p> <p>۷-۲. هزینه آموزش (چنانچه به کار بستن سیستم، مستلزم آموزش باشد)</p> <p>۸-۲. هزینه تجهیزات و اسباب و اثاثیه تازه برای مکان انجام پروژه (رایانه، میز، صندلی، و...)</p> <p>۹-۲. هزینه نرم‌افزار و پشتیبانی و هزینه منابع و پایگاه‌های اطلاعاتی الکترونیکی</p>	<p>(همتی و همکاران: ۱۳۹۳)، موخری و ری (۲۰۱۷)، بوس و همکاران (۲۰۱۴)، یانگ و همکاران (۲۰۱۲)، جواهری (۱۳۸۳)، پورفرج و همکاران (۱۳۹۱)، نجفی‌خواه و همکاران (۱۳۹۵)</p>
۳. زمان‌بندی	<p>۱-۳. زمان توسعه یک سیستم</p> <p>۲-۳. امکان شناور ساختن پروژه</p> <p>۳-۳. روشن کردن الزامی یا دلخواه بودن زمان تحویل پروژه</p>	<p>پاور و همکاران (۱۹۹۰)، موخری و ری (۲۰۱۷)، بوس و همکاران (۲۰۱۴)</p>
۴. عملیاتی	<p>۱-۴. پذیرا بودن بستر سازمان، تشکیلات، و محیط برای پیاده‌سازی پروژه</p> <p>۲-۴. همخوانی سازمانی و فرهنگی</p> <p>۳-۴. تأثیر سیستم پیشنهادی روی کارکنان و مدیران</p> <p>۴-۴. مناسب بودن طراحی محیط کار برای انجام پروژه</p> <p>۵-۴. وظایف مدیران و کاربران نهایی</p> <p>۶-۴. مقاومت کاربران نهایی</p> <p>۷-۴. احساس کاربران نهایی درباره انجام یک فرایند جدید</p> <p>۸-۴. تحت تأثیر قرار گرفتن محیط کاری و اندازه این تأثیر</p> <p>۹-۴. سود سیستم پیشنهادی برای سازمان</p>	<p>(شرف: ۱۳۸۶، نقل از سالاری: ۱۳۹۵)، موخری و ری (۲۰۱۷)، بوس و همکاران (۲۰۱۴)، ادیب‌منش و همکاران (۱۳۹۰)، شید (۲۰۱۳)</p>
۵. عوامل انسانی	<p>۱-۵. واکنش مردم به کاربرد یک سیستم جدید</p> <p>۲-۵. دیرکرد یا اشکال در ساخت سیستم پیشنهادی بر اثر واکنش مردم/ کاربران</p>	<p>پاور و همکاران (۱۹۹۰)</p>
۶. قانونی حقوقی	<p>۱-۶. امکان تناقض سیستم پیشنهادی با الزام‌های قانونی</p> <p>۲-۶. قوانین سرمایه‌گذاری خارجی</p> <p>۳-۶. قوانین اعطای حق مالکیت یا کنترل دارایی عمومی یا قوانین مسئولیت تحویل سرویس‌های عمومی به شخص ثالث</p> <p>۴-۶. چگونگی برخورد قانونی با منابع درآمد</p>	<p>(ویتنی: ۲۰۰۷، نقل از صرافی‌زاده و علیپور)، موخری و ری (۲۰۱۷)، بوس و همکاران (۲۰۱۴)، (ppp-certification.com)</p>

ابعاد	معیارها/ مقیاس‌ها	منابع
۷. فرهنگی	۱-۷. پذیرش فرهنگی سیستم پیشنهادی در جو سازمانی	(ویتنی: ۲۰۰۷، نقل از صرافی‌زاده و علیپور، ۱۳۹۰)
۸. منابع	۱-۷. منابع انجام پروژه ۲-۷. امکانات انجام پروژه ۳-۷. امکان استفاده از یارانه دولتی	(ppp-virtualservices.com) یانگ و همکاران (۲۰۱۲)، جواهری (۱۳۸۳)، نجفی‌خواه و همکاران (۱۳۹۵)
۹. بازاریابی	۱-۹. مشتریان محصول پس از ساخت آن ۲-۹. آزمایش محصول پیش از عرضه به بازار	جواهری (۱۳۸۳)، پورفرج و همکاران (۱۳۹۱)، (ppp-evirtualservices.com)

۲-۲- پژوهش‌های ساخت پیکره زبانی

واینه (۲۰۰۵) دیدگاه‌های پژوهشگران سرشناس ساخت پیکره و پیشنهادهای کاربردی آنان را برای ساخت پیکره نوشته است. راهنمایی‌های این پژوهشگران می‌تواند در ساخت پیکره زبانی کمک کنند. کار وی تعریف‌های پیکره، افزودن حاشیه‌نویسی زبانی، داده‌های به‌کار رفته و پیکره آوایی را در بر دارد. عاصی (۱۳۸۲) پس از اشاره به تاریخچه رشته زبان‌شناسی پیکره‌ای، به رویکردهای نظری و کاربردی زبان‌شناسی پیکره‌ای، گونه‌های پیکره‌های زبانی، نشانه‌گذاری پیکره‌ها، سازمان‌دهی پیکره‌ها، پایگاه‌های داده زبانی و نقش رایانه در آنها پرداخته است. سپس برخی از پیکره‌های کلیدی زبان انگلیسی و فارسی و نیز چند پروژه فرهنگ‌نگاری را معرفی کرده است که با بهره‌گیری از پیکره‌های زبانی انجام شده‌اند.

پیکره‌ها می‌توانند مانند پیکره بالینی روبرتز و همکاران (۲۰۰۹) تخصصی و دارای هدف‌های ویژه‌ای باشند. این پیکره حاشیه‌نویسی معنایی و متون حوزه بالینی را در بر دارد که برای ساخت و ارزیابی سیستم‌های خودکار بازیابی اطلاعات کلیدی بالینی بیماران به‌کار می‌رود. در این مقاله، چگونگی نمونه‌گیری از متون، روش حاشیه‌نویسی، توزیع حاشیه‌نویسی‌ها در پیکره و کاربرد پیکره در ساخت سیستم بازیابی اطلاعات آمده است. کلود توریدا (۲۰۱۶) ساخت گام به گام پیکره تخصصی^۱ و فهرست واژگان حاشیه‌نویسی شده را بر پایه فراوانی واژگان توضیح داده است. این گام‌ها؛ گزینش متون آموزشی، حذف واژگانی که قاموسی^۲ نیستند (مانند حرف اضافه، حرف ربط، جدول، شکل، و...)، واکاوی متن با نرم‌افزار «انت کانک»^۳، نگارش فهرست فراوانی واژه‌ها و نگارش فهرست واژگان حاشیه‌نویسی شده (تعیین مقوله نحوی^۴ واژگان در فهرست، افزودن تعریف واژگان، باهم‌آیی واژگان و نمونه‌ای از جمله‌ای که واژه در آن به‌کار رفته) است.

1. specialized corpus
2. Content words
3. AntConc
4. POS

پیکره‌ها را می‌توان از منابع گوناگون مانند کتاب‌ها، مجله‌ها، برگ‌های وب، و... ساخت. کامیابی گل و همکاران (۱۳۹۷) در مقاله خود، چگونگی ساخت پیکره مقاله‌های علمی پژوهشی را آورده‌اند. آنها برای ساخت پیکره این گام‌ها را آورده‌اند: طراحی و ساخت نرم‌افزار پیکره‌ساز، تعیین متغیرهای ساخت پیکره (کمینه شمار «توکن» فایل‌های یک متن برای آمدن در پیکره، گردآوری متون، تجزیه متون به جمله‌های تشکیل‌دهنده، بیرون آوردن ریشه واژه‌ها، برچسب‌گذاری نحوی و فراهم‌آوری نرم‌افزار جداگانه برای بیرون‌آوری اطلاعات آماری از پیکره است.

دوران و همکاران (۲۰۱۴) به فرایند ساخت پیکره‌های آوایی پرداخته‌اند. آنان پس از تعریف پیکره آوایی، درباره اجزای کلیدی ساخت چنین پیکره‌هایی مانند ذخیره پیکره، چگونگی به اشتراک گذاشتن و کاربرد دوباره پیکره، پرسش‌هایی درباره نمایندگی و اندازه پیکره، گزینش داده خام و حاشیه‌نویسی پیکره سخن گفته و در پایان، نظریه‌های ساخت پیکره را بررسی کرده‌اند. آنها در این کتاب به ابعاد و مؤلفه‌های ساخت پیکره پرداخته‌اند که چگونگی طراحی پیکره، گردآوری داده‌ها، حاشیه‌نویسی (روش حاشیه‌نویسی و سیستم‌های آن)، و چگونگی حاشیه‌نویسی آوایی خودکار پیکره‌های گفتاری را در بر دارد. آنان به کاربرد آماری پیکره و چگونگی آرشیو و انتشار آن نیز پرداخته‌اند.

بی‌جن‌خان و همکاران (۲۰۱۱) نیز چیزهایی را بررسی کرده‌اند که به آنها در ساخت یک منبع نوشتاری زبانی، با نام «پیکره» پرداخته شده است. مؤلفه‌های ساخت پیکره، گونه و اندازه متون به کار رفته در پیکره و حاشیه‌نویسی آن است. این پیکره دارای بیش از ۳۵ هزار فایل متنی است که هر کدام دارای متن کامل یا نمونه تصادفی از یک فایل متنی هستند. برای حاشیه‌نویسی فارسی در پیکره، از دستورکار «ایگلز»^۲ پیروی شده است. این دستورکار برای زبان‌های اروپایی و برخی از زبان‌های غیراروپایی مانند عربی هم به کار رفته است. آنها باور دارند که از این دستورکار برای فارسی نیز می‌توان بهره برد. دستورکار کاربرد این دستورکار، سلسله مراتبی از برچسب‌ها است. برای حاشیه‌نویسی نیز روش نیمه‌خودکار به کار رفته است. در ساخت این پیکره، به ساخت اضافه و هم‌نگاره‌ها نیز پرداخته شده است که در واکاوی متن دشواری‌هایی را پدید می‌آورند. یکی دیگر از گونه‌های پیکره، پیکره موازی است. پیکره‌های موازی از متونی درست شده‌اند که با ترجمه‌هایی از یک یا بیش از یک زبان دیگر در پیوند هستند. تابع بردار (۱۳۹۳) به ساخت پیکره زبانی موازی با کاربرد پیکره‌های قیاس‌پذیر پرداخته است. در این کار مؤلفه‌هایی مانند کاربرد پایگاه دانش «ویکی‌پدیا»، کاربرد شیوه‌مبته بر بازیابی اطلاعات (که بر پایه برچسب گروه و پیوند^۳ برای دسته‌بندی مقاله‌های همانند کار می‌کند)، دیدن ویژگی‌هایی مانند همانندی ترجمه و اندازه آن از دیدگاه ترازبندی برای امتیازدهی جمله‌ها و کاربرد مدل خطی برای وزن‌دهی بهینه به هر یک از این خصیصه‌هاست. قاسمی‌زاده و همکاران (۱۳۸۵) نیز پیکره فارسی موازی با شمار بسیاری از زبان‌های اروپایی را معرفی و نخستین گام‌ها را برای ساخت منابع بنیادین پردازش

1. token
2. EAGLES
3. link

زبان فارسی بیان کرده‌اند. این گام برای معرفی ویژگی‌های صرفی نحوی زبان فارسی و رمزنگاری آن‌ها بر پایه مدل «ایگلز/مولتکست»^۱ و منابع ویژه «مولتکست - ئیست»^۲ است. این مقاله پس از معرفی کوتاه زبان فارسی، با تأکید بر شیوه نگارش و ویژگی‌های صرفی نحوی آن، دسته‌بندی تازه‌ای برای مقوله‌های گفتاری فارسی و همچنین دستور نگارشی را برای فارسی در محیط دیجیتال ارائه می‌کند. ارائه ویژگی‌های صرفی نحوی زبان فارسی برابر خطمشی «ایگلز/مولتکست»، ارائه دسته‌بندی تازه از مقوله‌های گفتاری زبان فارسی، معرفی روش تازه برای ارائه متون الکترونیکی فارسی، و برچسب‌گذاری پیکره از ابعاد و مؤلفه‌های بررسی شده در ساخت چنین پیکره‌ای بودند.

از کوشش‌هایی که برای ساخت پیکره‌های دو زبانه فارسی‌انگلیسی انجام گرفته است، می‌توان از پیکره ساخت دشتبانی (۱۳۹۱) نام برد. وی از ترازبندی داده‌های پیکره، ساخت یک واژه‌نامه دوزبانه، کاربرد هوش مصنوعی، اندازه‌گیری امتیاز همانندی دو جمله و ذخیره اطلاعات نگاشت جمله‌های دو مجموعه انگلیسی و فارسی در پایگاه داده‌های پیکره همچون ابعاد و مؤلفه‌های برای ساخت چنین پیکره‌ای نام می‌برد.

از دیگر کاربردهای پیکره، فرهنگ‌نگاری و نگارش دستور زبان است. بی‌جن‌خان (۱۳۸۳) از نقش پیکره‌های زبانی در نگارش دستور زبان سخن گفته است. به باور وی اگر چه با تأکید زبان‌شناسان زایشی بر داده‌های بالقوه و نه بالفعل زبانی برای کشف پیچیدگی‌های ذهن انسان، جایگاه پیکره‌ها در واکاوی دستوری کاهش یافته، ولی با رشد شتابان فناوری اطلاعات، بایستگی ساخت پیکره‌ها و دادگان‌های زبانی با اندازه بسیار بزرگ از اولویت بالایی برخوردار شده است. وی پس از آوردن تعریف پیکره زبانی از کریستال (۱۹۹۴)، به بررسی نقدهایی می‌پردازد که بر پیکره‌های زبانی شده است. وی در پایان یک نرم‌افزار پیکره زبانی را معرفی و توصیف می‌کند. گام‌های تولید پیکره زبانی از دیدگاه وی (۱) گردآوری داده‌ها؛ (۲) آماده‌سازی داده‌ها؛ (۳) برچسب‌دهی نحوی معنایی؛ و (۴) آمارگان هستند.

با بررسی پژوهش‌های ساخت پیکره زبانی، شاخص‌های کلیدی ساخت پیکره زبانی به دست آمدند. این شاخص‌ها در ساخت هر پیکره‌ای، چه نوشتاری/متنی و چه گفتاری/آوایی به کار می‌روند. چکیده شاخص‌های کلیدی ساخت پیکره زبانی بر پایه پژوهش‌های سینکلر (۲۰۰۴)، واتام (۲۰۱۵)، اندرسون و همکاران (۲۰۰۷)، بایبر (۱۹۹۳)، آتکینز و همکاران (۱۹۹۲)، مک‌انری و همکاران (۲۰۰۶)، بیانچی (۲۰۱۲)، کاواگلیا (۲۰۰۲) و استون و برنارد (۱۹۹۷) در جدول دو آمده‌اند.

1. EAGLES/MULTEXT

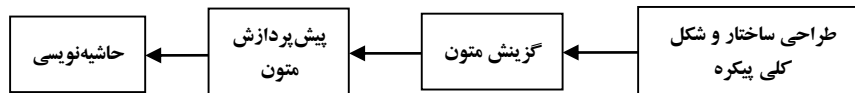
2. MULTEXT-East

جدول ۳: شاخص‌های کلیدی ساخت پیکره زبانی

شرح	شاخص
کنار نهادن متون بدون پیوند با هدف پیکره	یک‌دستی ^۱
عمومی ^۲ / تخصصی ^۳	گونه پیکره
شمار واژه‌های پیکره	اندازه ^۴
۱. چه زبانی و چه اندازه به کار رود ۲. متون چه اندازه تخصصی باشند ۳. متون چه اندازه رسمی یا غیررسمی باشند ۴. گزینش یک پیکره مبنا	توازن ^۵
۱. اندازه ژانرها و گونه‌های متون در پیکره ۲. چگونگی گزینش متون در پیوند با هر ژانر	نمایندگی ^۶
۱. گرایش / سوگیری سازندگان پیکره ۲. معیارهای گزینش نمونه‌ها: □ متن (گفتاری / نوشتاری / ...) □ واحد نمونه‌گیری (کتاب / مجله / مقاله / ...) □ زمینه متن (دانشگاهی / عمومی / ...) □ زبان متن (معیار / گونه ویژه زبانی) □ مکان (گونه زبانی وابسته به مکانی ویژه) □ زمان (تاریخ انتشار متن) ۳. روش نمونه‌گیری (تصادفی ساده / لایه لایه) ۴. اندازه نمونه (متن کامل / بریده متن) ۵. ژانر متون (ادبیات / کتاب درسی / قوانین و مقررات / ...)	نمونه‌گیری ^۷

نمودار یک فرایند و گام‌های ساخت پیکره متنی را بر پایه پژوهش‌های لیچ (۲۰۰۴)، برنارد (۲۰۰۷)، یاری (۱۳۹۴)، بی‌جن‌خان و همکاران (۲۰۱۱)، و مگردومیان (۲۰۰۴) نشان می‌دهد.

1. homogeneity
2. general corpus
3. specialized corpus
4. size
5. balance
6. representativeness
7. sampling



نمودار ۱: فرایند ساخت پیکره متنی

۳- روش‌شناسی پژوهش

رویکرد این پژوهش کیفی است. هنگامی این رویکرد به کار می‌رود که پژوهش در پی «ساخت»^۱ باشد نه «آزمون»^۲. اگر پژوهشی اکتشافی و به دنبال «ساخت» مفاهیم، مدل‌ها و چارچوب‌هایی باشد، این رویکرد به کار می‌رود (بومارد و ایبرت، ۲۰۰۱: ۷۹-۸۰). هدف از این پژوهش، ساخت مدل امکان‌سنجی ساخت پیکره متنی و بنابراین رویکرد آن نیز کیفی بود. این مدل بر پایه داده‌هایی طراحی شد که از بررسی پژوهش‌های پیشین و گردآوری دیدگاه متخصصان موضوعی به دست آمد.

گردآوری دیدگاه متخصصان موضوع درباره مدل امکان‌سنجی ساخت پیکره متنی با کاربرد روش دلفی^۳ (علیدوستی، ۱۳۸۵) و در دو دور انجام شد. کاربرد روش دلفی، ساخت دادن به فرایند ارتباطات میان افراد است به گونه‌ای که چنین فرایندی در فراهم کردن زمینه درگیری آنان به عنوان یک کل، با مسئله یا موضوعی پیچیده مؤثر باشد. این ارتباط ساخت یافته با بازخوران اطلاعات و دانش افراد، ارزیابی قضاوت یا نظر گروه، ارائه فرصتی به افراد برای بازنگری در نظرشان و تا حدی پنهان نگه داشتن پاسخ‌های افراد از یکدیگر تأمین می‌شود (لینستون و تروف، ۱۹۷۵، نقل از علیدوستی، ۱۳۸۵). روش دلفی در عمل، مجموعه‌ای از پرسشنامه‌ها یا دوره‌های^۴ پشت سر هم به همراه بازخوران کنترل شده‌ای است که تلاش دارد به اتفاق نظر میان یک گروه از افراد خبره^۵ درباره یک موضوع خاص دست یابد (لینستون و تروف، ۱۹۷۵؛ مک‌کنا، ۲۰۰۰، نقل از علیدوستی، ۱۳۸۵). روی هم رفته روش دلفی دارای چند گام کلیدی است (میتچل، ۱۹۹۱؛ جونز و هاتر، ۱۹۹۵؛ کلایتون، ۱۹۹۷؛ هاسون و همکاران، ۲۰۰۰؛ اکولیا و پالوسکی، ۲۰۰۴، نقل از علیدوستی، ۱۳۸۵): در گام نخست، مسئله پژوهش تعریف و بر این پایه، ویژگی‌های شرکت‌کنندگان در پل دلفی تعیین می‌شوند. سپس نامزدهای مشارکت در این پل شناسایی و از آنان دعوت به عمل می‌آید. این گام با تعیین اعضای پل به پایان می‌رسد. گام دوم روش دلفی به تولید ایده در زمینه مسئله پژوهش اختصاص دارد. در این گام، اعضای پل ایده‌های خود را درباره عوامل مرتبط با مسئله پژوهش ارائه می‌کنند. پژوهشگر با تحلیل و پالایش این ایده‌ها، حذف تکراری‌ها و کاربرد واژگان یکسان؛ فهرست نهایی عوامل مرتبط با مسئله پژوهش را استخراج می‌کند. در این گام ممکن است پژوهشگر، خود نیز عواملی را از پیش تعیین و آنها را برای نظرخواهی در اختیار اعضا قرار دهد.

1. constructing
2. test
3. Delphi method
4. rounds
5. expert panel

در گام دوم می‌توان از هر دو روش هم استفاده نمود، به این معنا که افزون بر دریافت ایده‌های اعضای پنل، دیدگاه آنها را نیز درباره عواملی که از پیش تعیین شده‌اند، درخواست کرد. فرایند تولید ایده‌های جدید از دور نخست روش دلفی آغاز می‌شود. در این دور از اعضای پنل خواسته می‌شود که آزادانه دیدگاه خود را در پاسخ به پرسش‌های مطرح شده، ارائه کنند. افزون بر این در دور نخست باید پرسشنامه با دقت نوشته شود تا از درستی نگارش و جمله‌بندی و قابلیت درک پرسش‌ها اطمینان حاصل گردد. تحلیل داده‌های پرسشنامه‌ها در بر دارنده تحلیل و مدیریت دقیق داده‌های کمی و کیفی است. داده‌های حاصل از دور نخست دلفی اغلب کیفی هستند و می‌توان آنها را با فنونی مانند تحلیل محتوا، تحلیل کرد. داده‌های این دور با گروه‌بندی اقلام مشابه در کنار هم تحلیل می‌شوند. در این گام واژه‌های گوناگونی که برای یک مقوله واحد به کار رفته‌اند، از سوی پژوهشگر برای ارائه یک توصیف واحد، در یک گروه قرار می‌گیرند. چنین توصیف‌ها و گروه‌بندی‌ها باید به‌گونه‌ای صورت پذیرند که از جمع‌بندی و ارائه صحیح داده‌ها اطمینان وجود داشته شود. در حقیقت، این گام برای کاهش تعداد عوامل به تعداد قابل قبول برای ادامه کار انجام می‌شود. گام بعدی به بازنگری در عوامل می‌پردازد. روش دلفی تا جایی ادامه می‌یابد که میان اعضا اتفاق نظر حاصل شود یا شمار شرکت‌کنندگان به کمتر از حد لازم برسد.

در این پژوهش، پس از تعریف مسئله پژوهش، ویژگی‌های اعضای پنل دلفی روشن شد. این اعضا ۱۰ نفر با تخصص در رشته‌های زبان‌شناسی همگانی/عمومی، زبان‌شناسی رایانشی، علوم ریاضی، علوم کامپیوتر، مهندسی کامپیوتر و اصطلاح‌نامه نگاری با دست‌کم سه سال پیشینه در این زمینه بودند. سپس ابعاد و مؤلفه‌های مدل مفهومی امکان‌سنجی ساخت پیکره‌های متنی و همچنین ابزارها و حاشیه‌نویسی‌های آن از بررسی پژوهش‌های پیشین به‌دست آمد و در یک پرسشنامه نوشته شد. این پرسشنامه را سه تن از متخصصان زبان‌شناسی رایانشی بررسی و تأیید کردند (پیوست ۱). سپس پرسشنامه به اعضای پنل داده و دیدگاه ایشان درباره این ابعاد و مؤلفه‌ها در یک پیوستار پنج گزینه‌ای (یک تا پنج) گردآوری و دیدگاه‌هایی پذیرفته شدند که دارای میانگین برابر یا بزرگ‌تر از چهار بودند. در این دور، دیدگاه‌های تازه اعضای پنل نیز گردآوری شد. همه دیدگاه‌ها تحلیل و پالایش و در نگارش یک‌دست شدند در گام دوم نیز دیدگاه اعضای پنل درباره دیدگاه‌های تازه گرفته و با همان سنجه بررسی شدند و در پایان ابعاد و مؤلفه‌های ساخت پیکره‌های متنی به‌دست آمد. درباره ابزارها و انواع حاشیه‌نویسی‌ها نیز در دور نخست، آنهایی که بیش از پنج پاسخ‌گو پذیرفته بودند، گزیده شدند و پیشنهادهای تازه پاسخ‌گویان نیز در مدل نهایی امکان‌سنجی ساخت پیکره متنی آورده شد. نمودار دو فرایند این پژوهش را نشان می‌دهد.



نمودار ۲: گام‌های پژوهش

۴- یافته‌ها

۴-۱- چارچوب مفهومی امکان‌سنجی ساخت پیکره متنی

بر پایه بررسی پژوهش‌های پیشین در زمینه ابعاد و مؤلفه‌های امکان‌سنجی و همچنین چگونگی ساخت پیکره‌های متنی، یک چارچوب مفهومی برای امکان‌سنجی ساخت پیکره متنی با هفت بُعد و ۲۸ مؤلفه به‌دست آمد (جدول ۴).

جدول ۴: ابعاد و مؤلفه‌های امکان‌سنجی ساخت پیکره متنی

۱. بُعد فنی
۱-۱. بررسی نیاز به زبان‌شناس همگانی / عمومی در تیم ساخت پیکره
۲-۱. بررسی نیاز به زبان‌شناس رایانشی در تیم ساخت پیکره
۳-۱. بررسی نیاز به برنامه‌نویس رایانه‌ای در تیم ساخت پیکره
۴-۱. بررسی کاربرد نرم‌افزارهای آماده در مقایسه با تولید نرم‌افزار جدید برای ساخت شکل کلی پیکره
۵-۱. بررسی نیاز به پشتیبانی برون‌سازمانی از پیکره
۶-۱. طراحی زیربنای پیکره (چارچوب و ساختار)، به عنوان گامی از ساخت پیکره
۷-۱. انتخاب نوع متون، به عنوان گامی از ساخت پیکره
۸-۱. تعیین حجم متون، به عنوان گامی از ساخت پیکره
۹-۱. پیش‌پردازش متون، به عنوان گامی از ساخت پیکره
۱۰-۱. حاشیه‌نویسی (برچسب‌گذاری و...)، به عنوان گامی از ساخت پیکره

<p>۲. بُعد اقتصادی</p> <p>۱-۲. برآورد هزینه متخصصان ساخت پیکره</p> <p>۲-۲. برآورد هزینه برنامه‌نویسی رایانه</p> <p>۳-۲. برآورد هزینه حاشیه‌نویسی دستی</p> <p>۴-۲. برآورد هزینه آموزش کاربران</p> <p>۵-۲. برآورد هزینه خرید نرم‌افزار</p> <p>۶-۲. برآورد هزینه وسایل و تجهیزات اداری (رایانه، میز، صندلی و...)</p>
<p>۳. بُعد زمان‌بندی</p> <p>۱-۳. برآورد زمان‌بندی ساخت پیکره</p>
<p>۴. بُعد قانونی / حقوقی</p> <p>۱-۴. بررسی الزام‌های قانونی و حقوقی در ساخت پیکره</p> <p>۲-۴. بررسی اجازه‌نامه کاربرد متون در ساخت پیکره</p> <p>۳-۴. بررسی حقوق مادی پیکره (پس از ساخت)</p>
<p>۵. بُعد عملیاتی</p> <p>۱-۵. بررسی پذیرش و آمادگی سازمان در کاربرد پیکره</p> <p>۲-۵. بررسی تأثیر پیکره بر سازوکارهای درون‌سازمانی</p> <p>۳-۵. بررسی دستاوردهای سازمان با ساخت پیکره</p> <p>۴-۵. پیش‌بینی وظایف کاربران در ساخت پیکره</p> <p>۵-۵. پیش‌بینی وظایف مدیران سازمان در ساخت پیکره</p>
<p>۶. بُعد تأمین منابع</p> <p>۱-۶. پیش‌بینی منابع تأمین مالی برای ساخت پیکره</p>
<p>۷. بُعد بازاریابی</p> <p>۱-۷. شناخت کاربران پیکره</p> <p>۲-۷. آزمون عملکرد پیکره پیش از عرضه به بازار</p>

افزون بر مؤلفه‌های امکان‌سنجی ساخت پیکره متنی، ابزارها و حاشیه‌نویسی‌ها در ساخت پیکره نیز از بررسی پژوهش‌های پیشین به‌دست آمدند. این یافته‌ها؛ ابزارهای جداسازی^۱، نرمال‌سازی^۲ یا پیش‌پردازش متن،

1. tokenizer
2. normalizer

«پارسر»^۱، ریشه‌یاب^۲، برچسب‌گذاری اجزای واژگانی کلام^۳، تشخیص‌دهنده گروه‌های نحوی^۴، و خطایاب^۵ و همچنین حاشیه‌نویسی‌های نحوی^۶، کاربردی^۷، گفتمانی^۸، سبکی^۹، و واژگانی^{۱۰} را در بر داشتند.

۴-۲- یافته‌های دور نخست دلفی

پرسشنامه دور نخست دلفی سه بخش داشت. در بخش نخست، دیدگاه پاسخ‌گویان درباره مؤلفه‌های ساخت پیکره متنی پرسیده شد. در بخش دوم نیز از پاسخ‌گویان درخواست شد که مؤلفه‌های کلیدی فنی، اقتصادی، حقوقی قانونی، زمان‌بندی، بازاریابی، عملیاتی و تأمین منابع را برای امکان‌سنجی ساخت پیکره متنی پیشنهاد کنند که در بخش نخست نبودند. در بخش سوم نیز دیدگاه پاسخ‌گویان درباره ابزارها و حاشیه‌نویسی‌های ساخت پیکره متنی پرسیده و در کنار آنها خواست شد که ابزارها و حاشیه‌نویسی‌های تازه را هم پیشنهاد کنند. این پرسشنامه را ۱۰ تن پاسخ دادند که میانگین پاسخ آنها به هر یک از مؤلفه‌های امکان‌سنجی در بخش نخست پرسشنامه در جدول پنج آمده است.

جدول ۵: میانگین پاسخ اعضای پانل دلفی در دور نخست به مؤلفه‌های امکان‌سنجی

ساخت پیکره متنی

میانگین پاسخ‌ها (از پنج)	شماره گویه	میانگین پاسخ‌ها (از پنج)	شماره گویه	میانگین پاسخ‌ها (از پنج)	شماره گویه	میانگین پاسخ‌ها (از پنج)	شماره گویه
۳ / ۸	۴-۱	۴ / ۵	۳-۱	۴ / ۵	۲-۱	۴ / ۷	۱-۱
۴ / ۹	۸-۱	۴ / ۹	۷-۱	۴ / ۴	۶-۱	۴ / ۰	۵-۱
۴ / ۹	۲-۲	۴ / ۸	۱-۲	۴ / ۴	۱۰-۱	۴ / ۶	۹-۱
۴ / ۴	۶-۲	۴ / ۶	۵-۲	۴ / ۱	۴-۲	۴ / ۵	۳-۲
۴ / ۹	۳-۴	۴ / ۷	۲-۴	۴ / ۴	۱-۴	۳ / ۷	۱-۳
۳ / ۶	۴-۵	۴ / ۲	۳-۵	۳ / ۷	۲-۵	۴ / ۳	۱-۵
۴ / ۵	۲-۷	۴ / ۲	۱-۷	۳ / ۷	۱-۶	۳ / ۸	۵-۵

1. parser
2. stemming
3. pos tagging
4. chunker
5. error detector
6. syntactic
7. pragmatic
8. discourse
9. stylistic
10. lexical

همان‌گونه که در جدول شش دیده می‌شود، چهار گویه «بررسی تأثیر پیکره بر سازوکارهای درون‌سازمانی»، «پیش‌بینی وظایف مدیران سازمان در ساخت پیکره»، «پیش‌بینی وظایف کاربران در ساخت پیکره»، و «بررسی کاربرد نرم‌افزارهای آماده در مقایسه با تولید نرم‌افزار جدید برای ساخت شکل کلی پیکره» دارای میانگینی کمتر از چهار هستند و از مؤلفه‌ها کنار گذاشته می‌شوند و دیگر گویه‌ها می‌مانند. در بخش دوم پرسشنامه، پاسخ‌گویان نه مؤلفه تازه را در سه بُعد امکان‌سنجی ساخت پیکره متنی پیشنهاد کردند که در جدول شش دیده می‌شوند.

جدول ۶: ابعاد و مؤلفه‌های تازه امکان‌سنجی ساخت پیکره متنی از دیدگاه اعضای پانل دلفی

۱. بُعد فنی
۱-۱. بررسی نیاز به رابط کاربر ^۱ برای کار کردن کاربران با پیکره
۲-۱. بررسی نیاز به همکاری پژوهشگرانی که تجربه ساخت پیکره دارند
۲. بُعد اقتصادی
۱-۲. برآورد هزینه اولیه تأمین زیرساخت (سرور، شبکه، و...)
۲-۲. برآورد هزینه نگهداری زیرساخت (سرور، شبکه، و...)
۳-۲. برآورد هزینه ساخت ابزار گردآوری متن
۴-۲. برآورد هزینه ساخت ابزار کار با داده‌های پیکره
۵-۲. تصمیم‌گیری درباره چگونگی دسترسی کاربران به پیکره (رایگان یا با پرداخت هزینه)
۶-۲. برآورد هزینه بروز کردن داده‌های پیکره در پیکره‌های پیش‌گر/بروزشونده
۳. بُعد زمان‌بندی
۱-۳. فازبندی زمان ساخت پیکره

پاسخ‌گویان در بخش سوم پرسشنامه، دیدگاه خود را درباره ابزارها و حاشیه‌نویسی‌ها نیز گفتند. برخی از پاسخ‌گویان باور داشتند که بسته به کاربرد پیکره (خلاصه‌سازی متن، استخراج کلیدواژه و استخراج نوع متن (مهندسی، علوم انسانی، و...)) می‌توان این ابزارها را به کار برد. بدین‌سان پاسخ‌گویان همه ابزارها و حاشیه‌نویسی‌ها را پذیرفتند و ابزارها و حاشیه‌نویسی‌های دیگری را نیز پیشنهاد کردند که در جدول هفت دیده می‌شوند.

جدول ۷: پیشنهادهای اعضای پانل درباره ابزارهای پردازش زبان و حاشیه‌نویسی

زمینه	پیشنهاد اعضای پانل در دور نخست
ابزارهای پردازش زبان	۱. حاشیه‌نویسی پیکره ۲. شناسایی موجودیت‌های نامدار ^۱ ۳. خزشگر ^۲
حاشیه‌نویسی	۱. فراداده‌گذاری (موضوع متن، جنسیت نویسنده، سال انتشار، و...) ۲. حاشیه‌نویسی (نشانه‌گذاری) معنایی ^۳ ۳. حاشیه‌نویسی گفتمانی (حرکت‌های بلاغی) ^۴ ۴. برچسب‌گذاری نقش‌های معنایی ^۵ ۵. حاشیه‌نویسی استعاره ^۶

۴-۳- یافته‌های دور دوم دلفی

پرسشنامه دور دوم دلفی دو بخش داشت. در بخش نخست، دیدگاه پاسخ‌گویان درباره مؤلفه‌های ساخت پیکره متنی پرسیده شد که اعضای پانل در دور نخست پیشنهاد کرده بودند. در بخش دوم نیز از پاسخ‌گویان خواسته شد که دیدگاه خود را درباره ابزارهای و حاشیه‌نویسی‌هایی بگویند که در دور نخست پیشنهاد شده بودند. این پرسشنامه را نیز ۱۰ تن پاسخ دادند که میانگین پاسخ آنها به هر یک از مؤلفه‌های امکان‌سنجی در بخش نخست پرسشنامه در جدول هشت آمده است.

جدول ۸: میانگین پاسخ اعضای پانل دلفی در دور دوم به مؤلفه‌های امکان‌سنجی ساخت پیکره متنی

شماره گویه	میانگین پاسخ‌ها (از پنج)	شماره گویه	میانگین پاسخ‌ها (از پنج)	شماره گویه	میانگین پاسخ‌ها (از پنج)
۱-۱	۴ / ۳	۲-۱	۴ / ۴	۳-۱	۴ / ۴
۲-۲	۴ / ۳	۳-۲	۴ / ۶	۴-۲	۴ / ۵
۵-۲	۴ / ۳	۶-۲	۴ / ۵	۱-۳	۴ / ۹

همان‌گونه که در جدول نه دیده می‌شود، همه مؤلفه‌ها دارای میانگینی بیشتر از چهار هستند و برای ساخت مدل امکان‌سنجی ساخت پیکره متنی به کار می‌روند.

1. named entity recognition
2. crawler
3. semantic annotation
4. rhetorical moves
5. semantic role labeling
6. metaphor annotation

۵- بحث و نتیجه‌گیری

از آنجایی که بررسی پژوهش‌های پیشین مدل فراگیری را برای امکان‌سنجی ساخت پیکره زبانی نشان نداد، مدل به‌دست آمده در این پژوهش کلیدی‌ترین یافته آن است. این مدل با آمیختن یافته‌های دور نخست و دوم دلفی، دارای هفت بُعد و ۳۳ مؤلفه است (جدول ۹). که تا اندازه‌ای مانند ابعاد یافت شده در پژوهش‌های موخری و ری (۲۰۱۷) و بوس و همکاران (۲۰۱۴) است که پنج بُعد فنی، اقتصادی، قانونی حقوقی، عملیاتی، و زمان‌بندی را داشتند. مهارت‌ها و تخصص‌های فنی، طراحی زیرساخت، سخت‌افزار و نرم‌افزارها و مانند آنها برای انجام پروژه که از بررسی مؤلفه‌های امکان‌سنجی پژوهش‌های پیشین (مانند صرافی‌زاده و علیپور، ۱۳۹۰؛ کرنقان، ۲۰۱۲) به‌دست آمده‌اند، در طراحی مدل امکان‌سنجی ساخت پیکره به‌کار رفتند. افزون بر ابعاد و مؤلفه‌های آمده در پژوهش‌های پیشین (جدول ۲)؛ در بُعد فنی، گام‌هایی مانند انتخاب نوع متون، تعیین حجم متون، پیش‌پردازش متون و حاشیه‌نویسی متون که در فرایند ساخت پیکره متنی باید برداشته شوند، از پژوهش‌های پیشین (لیچ، ۲۰۰۴؛ برنارد، ۲۰۰۷؛ یاری، ۱۳۹۴؛ بی‌جن‌خان و همکاران، ۲۰۱۱ و مگردومیان، ۲۰۰۴) به‌دست آمده‌اند. در بُعد اقتصادی مؤلفه‌هایی مانند هزینه نیروی انسانی، نرم‌افزارها، آموزش کاربران، مشاوره با متخصصان، و مانند آنها آمدند که در پژوهش‌های همتی و همکاران (۱۳۹۳) و نجفی‌خواه و همکاران (۱۳۹۵) دیده شده‌اند. برخی از مؤلفه‌های یافت شده در این پژوهش مانند برآورد هزینه‌های تأمین زیرساخت، پشتیبانی، و ساخت ابزار گردآوری متن در پژوهش‌های پیشین دیده نمی‌شوند، چرا که با پروژه ساخت پیکره سروکار دارند و تخصصی هستند.

همان‌گونه که جدول ۸ و نمودار سه نشان می‌دهند، در امکان‌سنجی ساخت پیکره متنی باید به هفت بُعد و ۳۳ مؤلفه پرداخت. این ابعاد؛ فنی، اقتصادی، زمان‌بندی، قانونی حقوقی، عملیاتی، تأمین مالی و بازاریابی هستند. در هر بُعد نیز مؤلفه‌هایی را در بر دارد. در بُعد فنی، نیاز به نیروی متخصص در تیم ساخت پیکره مانند زبان‌شناس همگانی/عمومی، زبان‌شناس رایانشی، برنامه‌نویس رایانه‌ای، پژوهشگرانی که تجربه ساخت پیکره دارند، بررسی می‌شود. بررسی نیاز به پشتیبانی برون‌سازمانی از پیکره و رابط کاربر (User interface) برای کار کردن کاربران با پیکره از دیگر ویژگی‌های فنی ساخت پیکره است. در بُعد فنی گام‌هایی هم برای ساخت پیکره دیده می‌شوند که انتخاب نوع متون، تعیین حجم متون، پیش‌پردازش متون و حاشیه‌نویسی (برچسب‌گذاری و...) هستند. در بُعد اقتصادی باید برای مشاوره گرفتن و به‌کارگیری نیروهای متخصص، برآورد هزینه انجام شود. هزینه برنامه‌نویسی رایانه، حاشیه‌نویسی دستی، آموزش کاربران، خرید نرم‌افزار، خرید وسایل و تجهیزات اداری (رایانه، میز، صندلی و...)، تأمین زیرساخت (سرور، شبکه، و...)، نگهداری زیرساخت (سرور، شبکه، و...)، ساخت ابزار گردآوری متن، ساخت ابزار کار با داده‌های پیکره، و بروز کردن داده‌های پیکره در پیکره‌های پیش‌گر/به‌روزشونده نیز باید برآورد هزینه شود. مؤلفه دیگری که در بُعد اقتصادی دیده می‌شود، تصمیم‌گیری درباره چگونگی دسترسی کاربران به پیکره (رایگان یا با پرداخت هزینه) است. در بُعد زمان‌بندی، باید زمان‌بندی پروژه ساخت پیکره و گام‌های آن روشن شود. مؤلفه‌هایی که در بُعد حقوقی قانونی بررسی می‌شوند؛ الزام‌های قانونی و حقوقی و اجازه‌نامه کاربرد متون و حقوق مادی پیکره (پس از ساخت) هستند تا

ساخت پیکره از دیدگاه حقوقی قانونی با مانعی برخورد نکند. تأمین مالی بعد دیگری از امکان‌سنجی ساخت پیکره است که در آن باید منابع تأمین مالی برای ساخت پیکره پیش‌بینی شود. در بعد بازاریابی نیز شناخت کاربران پیکره و آزمون عملکرد پیکره، پیش از عرضه آن به بازار باید بررسی شوند.

- ابزارها و حاشیه‌نویسی

ابزارها و حاشیه‌نویسی‌های گوناگونی نیز می‌توانند در پیکره متنی به کار روند: ابزار جداسازی، ابزار نرمال‌سازی متن / ابزار پیش‌پردازش متن، ابزار «پارسر»، ابزار ریشه‌یاب، ابزاری برای برچسب‌گذاری اجزای واژگانی کلام، ابزار تشخیص‌دهنده گروه‌های نحوی، ابزار خطایاب، حاشیه‌نویسی نحوی، کاربردی، گفتمانی، سبکی، و واژگانی، ابزار حاشیه‌نویسی پیکره، ابزار شناسایی موجودیت‌های نامدار (named entity recognition)، خزش‌گر (crawler)، تعیین فراداده (موضوع متن، جنسیت نویسنده، سال انتشار و...)، حاشیه‌نویسی (نشانه‌گذاری) معنایی (semantic annotation)، در مورد حاشیه‌نویسی گفتمانی، مشخصاً حاشیه‌نویسی حرکت‌های بلاغی (rhetorical moves) چکیده‌ها، برچسب‌گذاری نقش‌های معنایی (semantic role labeling) و حاشیه‌نویسی استعاره (metaphor annotation)، باشد.

جدول ۹: ابعاد و مؤلفه‌های امکان‌سنجی ساخت پیکره متنی

مؤلفه‌ها	ابعاد
۱-۱. بررسی نیاز به زبان‌شناس همگانی / عمومی در تیم ساخت پیکره	۱. فنی
۲-۱. بررسی نیاز به زبان‌شناس رایانشی در تیم ساخت پیکره	
۳-۱. بررسی نیاز به برنامه‌نویس رایانه‌ای در تیم ساخت پیکره	
۴-۱. بررسی نیاز به پشتیبانی برون‌سازمانی از پیکره	
۵-۱. طراحی زیربنای پیکره (چارچوب و ساختار)، به‌عنوان گامی از ساخت پیکره	
۶-۱. انتخاب نوع متون، به‌عنوان گامی از ساخت پیکره	
۷-۱. تعیین حجم متون، به‌عنوان گامی از ساخت پیکره	
۸-۱. پیش‌پردازش متون، به‌عنوان گامی از ساخت پیکره	
۹-۱. حاشیه‌نویسی (برچسب‌گذاری و...)، به‌عنوان گامی از ساخت پیکره	
۱۰-۱. بررسی نیاز به رابط کاربری برای کار کردن کاربران با پیکره	
۱۱-۱. بررسی نیاز به همکاری پژوهشگرانی که تجربه ساخت پیکره دارند	
۱-۲. برآورد هزینه متخصصان ساخت پیکره	۲. اقتصادی
۲-۲. برآورد هزینه برنامه‌نویسی رایانه	
۳-۲. برآورد هزینه حاشیه‌نویسی دستی	
۴-۲. برآورد هزینه آموزش کاربران	
۵-۲. برآورد هزینه خرید نرم‌افزار	

۶-۲	برآورد هزینه وسایل و تجهیزات اداری (رایانه، میز، صندلی و...)	
۷-۲	برآورد هزینه اولیه تأمین زیرساخت (سرور، شبکه، و...)	
۸-۲	برآورد هزینه نگهداری زیرساخت (سرور، شبکه، و...)	
۹-۲	برآورد هزینه ساخت ابزار گردآوری متن	
۱۰-۲	برآورد هزینه ساخت ابزار کار با داده‌های پیکره	
۱۱-۲	تصمیم‌گیری درباره چگونگی دسترسی کاربران به پیکره (رایگان یا با پرداخت هزینه)	
۱۲-۲	برآورد هزینه بروز کردن داده‌های پیکره در پیکره‌های پیش‌گر / بروزشونده	
۱-۳	برآورد زمان‌بندی ساخت پیکره	۳. زمان‌بندی
۲-۳	فازبندی زمان ساخت پیکره	
۱-۴	بررسی الزام‌های قانونی و حقوقی در ساخت پیکره	۴. قانونی حقوقی
۲-۴	بررسی اجازه‌نامه کاربرد متون در ساخت پیکره	
۳-۴	بررسی حقوق مادی پیکره (پس از ساخت)	
۱-۵	بررسی پذیرش و آمادگی سازمان در کاربرد پیکره	۵. عملیاتی
۲-۵	بررسی دستاوردهای سازمان با ساخت پیکره	
۱-۶	پیش‌بینی منابع تأمین مالی برای ساخت پیکره	۶. تأمین مالی
۱-۷	شناخت کاربران پیکره	۷. بازاریابی
۲-۷	آزمون عملکرد پیکره پیش از عرضه به بازار	



نمودار ۳: مدل امکان‌سنجی ساخت پیکره متنی

در ساخت پیکره متنی، ابزارها و حاشیه‌نویسی‌های گوناگونی به کار می‌روند. ابزارهای جداسازی، نرمال‌سازی یا پیش‌پردازش متن، «پارسر»، ریشه‌یاب، برچسب‌گذاری اجزای واژگانی کلام، تشخیص‌دهنده گروه‌های نحوی و خطایاب از این شمار هستند. همچنین حاشیه‌نویسی نحوی، کاربردی، گفتمانی، سبکی و واژگانی، فراداده‌گذاری (موضوع متن، جنسیت پدیدآور، سال انتشار، و...)، معنایی، گفتمانی (حرکت‌های بلاغی)، برچسب‌گذاری نقش‌های معنایی و استعاره از حاشیه‌نویسی‌هایی هستند که در ساخت پیکره به کار می‌روند. از دیگر ابزارهای کاربردی در پیکره می‌توان ابزار حاشیه‌نویسی پیکره، ابزار شناسایی موجودیت‌های نامدار، و ابزار خزش‌گر را نام برد.

۶- پیشنهادها

پیشنهاد می‌شود پژوهشگران و سازمان‌های تأمین‌کننده مالی، پیش از آغاز هر پروژه ساخت پیکره متنی، نخست امکان‌سنجی آن را به انجام دهند تا بتوانند با شناخت و آگاهی بیشتری آن را به سرانجام برسانند. این شناخت و آگاهی، به‌ویژه در پروژه‌های بزرگ می‌تواند اثربخشی و کارایی را در پروژه به ارمغان آورد. در ساخت پیکره‌های بزرگ برای زبان فارسی، پیشنهاد می‌شود که بر پایه ابعاد و مؤلفه‌های امکان‌سنجی، نخست پیکره‌ای کوچک ساخته شود؛ سپس اندازه داده‌ها افزایش یابد. افزون بر این، در زبان فارسی نیاز به ساخت ابزارهایی نیز هست که با کاربرد آنها بتوان پیکره‌های بزرگ فارسی را سامان داد.

پیشنهاد می‌شود این مدل با روش اقدام‌پژوهی^۱ در یک پروژه ساخت پیکره متنی به کار رود تا برای آن، اعتبار تجربی نیز فراهم شود. پژوهش‌های کاربردی برای ساخت یا بهبود عملکرد ابزارهایی مانند خطایاب، تشخیص‌دهنده گروه‌های نحوی، پارسر، ابزار نرمال‌سازی متن و... و انواع حاشیه‌نویسی گفتمانی، کاربردی، سبکی، نحوی و... برای زبان فارسی از دیگر کارهایی است که نیاز به انجام آنها هست. برخی از مؤلفه‌های مدل می‌توانند جداگانه موضوع پژوهش باشند. برای نمونه «الزام‌های قانونی و حقوقی در ساخت پیکره» در بُعد قانونی-حقوقی یا «پذیرش و آمادگی سازمان در کاربرد پیکره» در بُعد عملیاتی، می‌توانند در پژوهش‌هایی جداگانه بررسی و راهکارها و مدل‌های آنها بررسی و تبیین شوند.

منابع

- ادیب‌منش، مرزبان؛ عسگری، محمدعلی و موسی‌پور، نعمت‌الله (۱۳۹۰). «امکان‌سنجی اجرای برنامه‌ریزی درسی مدرسه محور در دبیرستان‌های شهر کرمانشاه»، *پژوهش‌های برنامه درسی*، ۱(۲): ۸۱-۱۰۴.
- بی‌جن‌خان، محمود (۱۳۸۳). «نقش پیکره‌های زبانی در نوشتن دستور زبان: معرفی یک نرم‌افزار رایانه‌ای»، *مجله زبان‌شناسی*، ۱۹(۲): ۴۸-۶۷.
- پورفرج، اکبر؛ کرمی، مهرداد؛ نکویی، زهرا و طالب‌بیدختی، زهرا (۱۳۹۱). «امکان‌سنجی فرصت‌های سرمایه‌گذاری صنعت گردشگری. مطالعه موردی: شهرک تفریحی - توریستی طارم استان زنجان»، *فصل‌نامه مطالعات مدیریت صنعتی*، ۹(۲۵): ۲۱۴-۱۸۹.
- تابع‌دربار، علیرضا (۱۳۹۳). *ایجاد پیکره زبانی موازی به وسیله پیکره‌های قیاس‌پذیر*، کارشناسی ارشد. شیراز: دانشگاه شیراز.
- جواهری، حسن (۱۳۸۳). *مطالعات امکان‌سنجی پروژه‌های تفریحی و گردشگری*، اولین کنفرانس بین‌المللی مدیریت پروژه، تهران: گروه پژوهشی آریانا.
- دشتبانی، شکوفه (۱۳۹۱). *ساخت پیکره متنی فارسی حوزه فاوا*، دانشگاه بوعلی سینا، دانشکده فنی و مهندسی.

- سالاری، سمیه (۱۳۹۵). *امکان‌سنجی پیاده‌سازی منتورینگ در صلاحیت حرفه‌ای معلمان دوره ابتدایی*، تهران: دانشگاه علامه طباطبائی، دانشکده علوم تربیتی و روانشناسی.
- صراف‌زاده، اصغر و علیپور، وحیده (۱۳۹۰). «امکان‌سنجی ارائه محصول بیمه‌ای موجود به صورت الکترونیک در شرکت‌های بیمه پارسین، آسیا، البرز، رازی، توسعه، سینا و معلم»، *فصل‌نامه مدیریت توسعه و تحول*، ۷: ۳۹-۴۸.
- عاصی، مصطفی (۱۳۸۴). «گزارش کوتاهی از شکل‌گیری پایگاه داده‌های زبان فارسی در اینترنت»، *مجله پژوهشگران*، شماره ۲: ۱۳. قابل دسترسی در: <https://hawzah.net/fa/Magazine/View/6280/6282/69701>
- عاصی، مصطفی (۱۳۸۲). *از پیکره زبانی تا زبان‌شناسی پیکره‌ای*، مجموعه مقالات پنجمین کنفرانس زبان‌شناسی، دانشگاه علامه طباطبائی.
- علیدوستی، سیروس (۱۳۸۵). «روش دلفی: مبانی، مراحل و نمونه‌هایی از کاربرد»، *مدیریت و توسعه*، ۸ (۳۱): ۲۳-۸.
- قاسمی‌زاده، بهرنگ؛ رحیمی، سعید؛ سلاریان، مرتضی و بهاری‌سلیم، علی (۱۳۸۵). *گزارشی از ساخت نخستین پیکره چندزبانه برای زبان فارسی*، دوازدهمین کنفرانس سالانه انجمن کامپیوتر ایران، تهران: دانشگاه شهید بهشتی.
- کامیابی‌گل، عطیه؛ اخلاقی باقوجری، الهام؛ عسگریان، احسان و حبیبی، هانیه (۱۳۹۷). «استخراج اطلاعات از پیکره زبانی: معرفی پیکره مقالات علمی-پژوهشی دانشگاه فردوسی مشهد»، *مقالات آماده انتشار، فصل‌نامه کتابداری و اطلاع‌رسانی*، سازمان کتابخانه‌ها، موزه‌ها و مرکز اسناد آستان قدس رضوی.
- کوهستانی، منوچهر (۱۳۸۹). *بررسی خطاهای املائی و نگارشی در وبلاگ‌های فارسی و ماهیت زبان‌شناختی آنها*، کارشناسی ارشد، تهران، دانشگاه تهران.
- مرادی، مهدی و بحرانی، محمد (۱۳۹۴). «تشخیص خودکار جنسیت نویسنده در متون فارسی»، *پردازش علائم و داده‌ها*، ۱۲ (۴): ۸۳-۹۴.
- میرزایی، آزاده و مولودی، امیرسعید (۱۳۹۳). «نخستین پیکره نقش‌های معنایی در زبان فارسی»، *علم زبان*، ۲ (۳): ۲۹-۴۸.
- نجفی‌خواه، رقیه؛ جلالی دیزجی، علی و مومنی، عصمت (۱۳۹۵). «امکان‌سنجی ارائه خدمات مرجع مجازی در کتابخانه‌های عمومی استان گیلان»، *فصل‌نامه مطالعات دانش‌شناسی*، ۳ (۹): ۶۱-۸۰.
- همتی، عبدالناصر؛ مهدوی کلیشمی، قدیر و یاری، حمید (۱۳۹۳). «امکان‌سنجی ارائه بیمه عمر متصل به سهام در بازار بیمه کشور از دیدگاه خبرگان»، *دوفصل‌نامه جستارهای اقتصادی ایران*، ۱۱ (۲۲): ۹۵-۱۱۶.
- یاری، علیرضا (۱۳۹۴). *بررسی پیکره‌ها و ابزارهای پردازش زبان فارسی*، پژوهشگاه ارتباطات و فناوری اطلاعات، پژوهشکده ارتباطات و فناوری اطلاعات.
- AleAhmad, A; Amiri, H.; Darrudi, E.; Rahgozar, M; & Oroumchian, F. (2009). "Hamshahri: A Standard Persian Text Collection", *Knowledge-Based Systems, Dubai*, 22(5): 382-387. Elsevier.
- Al-Sulaiti, L; & Atwell, E. (2006). "The design of a corpus of contemporary Arabic", *International journal of corpus linguistics*, 11(2): 135-171. John Benjamins publishing company.

- Anderson, J; Beavan, D; & Kay, Ch. (2007). *The Scottish corpus of texts and speech*. In: *Beal, J.C., Corrigan, K.P. and Moisl, H.L. (eds.) Creating and Digitizing Language Corpora*, 17-34. Palgrave, New York, USA.
- Aston, G; & Burnard, L. (1997). *The NBC handbook exploring the British National Corpus with SARA*, Edinburgh University Press.
- Atkins, S; Clear, J; & Ostler, N. (1992). "Corpus design criteria". *Literary and Linguistic Computing*, 7 (1): 1-16. Oxford University Press (OUP).
- Baumard, Ph; and Jerome Ibert. (2001). *What approach with which data*. In *Doing Management Research: A Comprehensive Guide*, London, Sage.
- Baus, K; Radimersky, A; Iwanicki, M; & Albers, A. (2014). "Feasibility studies in the product development process", 24th CIRP design conference, *Procedia CIRP*, 21: 473-478. Elsevier.
- Bianchi, F. (2012). *Culture corpora and semantics: methodological issues in using elicited and corpus data for cultural comparison, Chapter 3: corpora and corpus linguistics*, University of Salento.
- Biber, D. (1993). "Representativeness in corpus design", *Literary and linguistic computing*, 8(4): 243-257. Oxford University Press (OUP).
- Bijankhan, M; Sheykhzadegan, J; Bahrani, M; & Ghayoomi, M. (2011). "Lesson from building a Persian written corpus: Peykare", *Language resources and evolution*, 45(2): 143-164. Springer.
- Burnard, L. (2007). *Reference guide for the British National Corpus. Research technologies service*, Oxford University Computing Services.
- Cavaglia, G. (2002). "Measuring corpus homogeneity using a range of measures for inter-document distance", *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*: 426-431. European Language Resources Association (ELRA). Spain.
- Claude Toriida, M. (2016). "Steps for creating specialized corpus and developing an annotated frequency-based vocabulary list", *TESL Canada journal/ revue TESL du Canada*, 34 (11): 87-105.
- Dehghani, M; Shakery, A; Asadpour, M; & Koushkestani, A. (2013). "A learning approach for email conversation thread reconstruction", *Journal of Information Science (JIS)*, 39(6): 846-863.
- Durand, J; Gut, U; & Gjert K. (2014). *The handbook of corpus phonology*, Oxford University Press (OUP).
- Fridrichova Klimova, Blanka (2014). "Using corpus linguistics in the development of writing", *Procedia- Social and behavioral sciences*, 141 (25): 124-128. Elsevier.
- <http://farsnet.nlp.sbu.ac.ir>
- <https://www.peykaregan.ir>
- Kernaghan, S. (2012). *Technical feasibility assessment*, ISET-international.
- Laurence, A. (2017). *Vocabulary learning and instruction*, 6 (2).

- Leech, G. (2004). *Developing Linguistic Corpora: a Guide to Good Practice. adding linguistic annotation. Edited by Martin Wynne*, ahds.literature, languages and linguistics, The Oxford Text Archive.
- McEnery, T; & Wilson, A. (2001). *Corpus Linguistics: An Introduction*, Edinburgh University Press.
- McEnery, T; Xiao, R; & Tono, Y. (2006). *Corpus-based language studies: and advanced resource book*, Routledge, London and New York.
- Megerdooian, K. (2004). "Developing a Persian part-of-speech tagger", Proceedings of the 1st workshop on Persian language and computer: 99-105.
- Mukherjee, M; & Roy, S. (2017). "Feasibility studies and important aspect of project management", *International journal of advanced engineering and management*, 2 (4): 98-100.
- Powers, M. J; Cheney, P. H; & Crow, G. (1990). *Structure system development. Analysis, design, implementation*, Boyd and Fraser publishing company.
- Roberts, A; Gaizauskas, R; Hepple, M.; Demetriou, G; Guo, Y; Roberts, I; & Setzer, A. (2009). "Building a semantically annotated corpus of clinical texts", *Journal of Biomedical Informatics*, 42(5): 950-966.
- Scheid, J. (2013). *An Example of Operational Feasibility. edited by: Michele McDonough*, Bright Hub Inc. at: <https://www.brighthubpm.com/project-planning/72872-an-example-of-operational-feasibility>
- Shamsfard, M; Hesabi, A; Fadaei, H; Mansoori, N; Famian, A; Bagherbeigi, S; Fekri, E. et al. (2010). "Semi Automatic Development of Farsnet; the Persian Wordnet", Proceedings of 5th Global WordNet Conference (GWA), Mumbai, India.
- Sinclair, J. (2004). *Developing Linguistic Corpora: a Guide to Good Practice. Chapter 1: Corpus and Text - Basic Principles*, Edited by Martin Wynne, ahds, literature, languages and linguistics, The Oxford Text Archive.
- Teubert, W. (2005). "My version of Corpus linguistics", *International journal of Corpus linguistics*, 10 (1): 1-13.
- Wattam, S. M. (2015). *Technological Advances in Corpus Sampling methodology*, Mathematics and Statistics School of Computing and Communications Lancaster University.
- Wayne, M. (2005). *Developing linguistic corpora: a guide to good practice*, Oxbow books, Literary and linguistic computing, 22 (1).
- Yang, J; Chen, W; Chen, B; & Jia, Y. (2012). "Economic feasibility analysis of a renewable energy project in the rural China", The 18th Biennial conference of International society for ecological modeling, Procedia environmental sciences 13: 2280-2283. Elsevier.

پوست ۱. پرسش‌نامه دور اول دلفی برای امکان‌سنجی ساخت پیکره

بخش نخست: دیدگاه شما درباره مؤلفه‌های امکان‌سنجی ساخت پیکره متنی که از بررسی پژوهش‌های پیشین به دست آمده‌اند.					
کاملاً مخالفم ۱	تا اندازه‌ای مخالفم ۲	مطمئن نیستم ۳	تا اندازه‌ای موافقم ۴	کاملاً موافقم ۵	با بودن هر یک از مؤلفه‌های زیر در امکان‌سنجی ساخت پیکره متنی چه اندازه موافق هستید؟
۱. بُعد فنی					
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۱-۱. بررسی نیاز به زبان‌شناس همگانی / عمومی در تیم ساخت پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۲-۱. بررسی نیاز به زبان‌شناس رایانشی در تیم ساخت پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۳-۱. بررسی نیاز به برنامه‌نویس رایانه‌ای در تیم ساخت پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۴-۱. بررسی کاربرد نرم‌افزارهای آماده در مقایسه با تولید نرم‌افزار جدید برای ساخت شکل کلی پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۵-۱. بررسی نیاز به پشتیبانی برون‌سازمانی از پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۶-۱. طراحی زیربنای پیکره (چارچوب و ساختار)، به عنوان گامی از ساخت پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۷-۱. انتخاب نوع متون، به عنوان گامی از ساخت پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۸-۱. تعیین حجم متون، به عنوان گامی از ساخت پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۹-۱. پیش‌پردازش متون، به عنوان گامی از ساخت پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۱۰-۱. حاشیه‌نویسی (برچسب‌گذاری و...)، به عنوان گامی از ساخت پیکره
۲- بُعد اقتصادی					
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۱-۲. برآورد هزینه متخصصان ساخت پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۲-۲. برآورد هزینه برنامه‌نویسی رایانه
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۳-۲. برآورد هزینه حاشیه‌نویسی دستی
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۴-۲. برآورد هزینه آموزش کاربران
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۵-۲. برآورد هزینه خرید نرم‌افزار
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۶-۲. برآورد هزینه وسایل و تجهیزات اداری (رایانه، میز، صندلی و...)
۳- بُعد زمان‌بندی					
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۱-۳. برآورد زمان‌بندی ساخت پیکره

بخش نخست: دیدگاه شما درباره مؤلفه‌های امکان‌سنجی ساخت پیکره متنی که از بررسی پژوهش‌های پیشین به دست آمده‌اند.					
کاملاً مخالفم ۱	تا اندازه‌ای مخالفم ۲	مطمئن نیستم ۳	تا اندازه‌ای موافقم ۴	کاملاً موافقم ۵	با بودن هر یک از مؤلفه‌های زیر در امکان‌سنجی ساخت پیکره متنی چه اندازه موافق هستید؟
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۴- بعد قانونی / حقوقی
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۴-۱. بررسی الزام‌های قانونی و حقوقی در ساخت پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۴-۲. بررسی اجازه‌نامه کاربرد متون در ساخت پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۴-۳. بررسی حقوق مادی پیکره (پس از ساخت)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۵- بعد عملیاتی
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۵-۱. بررسی پذیرش و آمادگی سازمان در کاربرد پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۵-۲. بررسی تأثیر پیکره بر سازوکارهای درون‌سازمانی
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۵-۳. بررسی دستاوردهای سازمان با ساخت پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۵-۴. پیش‌بینی وظایف کاربران در ساخت پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۵-۵. پیش‌بینی وظایف مدیران سازمان در ساخت پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۶- بعد تأمین منابع
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۶-۱. پیش‌بینی منابع تأمین مالی برای ساخت پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۷- بعد بازاریابی
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۷-۱. شناخت کاربران پیکره
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	۷-۲. آزمون عملکرد پیکره پیش از عرضه به بازار

بخش دوم: پیشنهادهای شما درباره مؤلفه‌های کلیدی امکان‌سنجی ساخت پیکره متنی که در بخش نخست نبودند.

<input type="checkbox"/>	۱. بعد فنی
<input type="checkbox"/>	۲. بعد اقتصادی
<input type="checkbox"/>	۳. بعد قانونی / حقوقی
<input type="checkbox"/>	۴. بعد زمان‌بندی
<input type="checkbox"/>	۵. بعد عملیاتی
<input type="checkbox"/>	۶. بعد تأمین منابع
<input type="checkbox"/>	۷. بعد بازاریابی

بخش سوم: دیدگاه شما درباره ابزارهای لازم برای ساخت پیکره متن

۱. به جز ابزارهای پردازش زبان طبیعی زیر، روی هم‌رفته چه ابزارهای دیگری را در ساخت پیکره متنی دارای کاربرد می‌دانید؟

(می‌توانید چند گزینه را نیز همزمان انتخاب کنید.)

- ♦ ابزار جداساز (Tokenizer)
 - ♦ ابزار نرمال‌سازی متن / پیش‌پردازش متن (Normalizer)
 - ♦ پارسر (Parser)
 - ♦ ابزار ریشه‌یابی (Stemming)
 - ♦ ابزار برچسب‌گذاری اجزای واژگانی کلام (POS tagging)
 - ♦ ابزار تشخیص‌دهنده گروه‌های نحوی (Chunker)
 - ♦ خطایاب (Error detector)
- خواهشمند است نام فارسی و انگلیسی ابزارهای دیگر را بنویسید:

۲. به جز حاشیه‌نویسی‌های زیر، روی هم‌رفته چه حاشیه‌نویسی‌هایی را در ساخت پیکره متنی دارای کاربرد می‌دانید؟

(می‌توانید چند گزینه را نیز همزمان انتخاب کنید.)

- ♦ حاشیه‌نویسی نحوی (Syntactic annotation)
 - ♦ حاشیه‌نویسی کاربردی (Pragmatic annotation)
 - ♦ حاشیه‌نویسی گفتمانی (Discourse annotation)
 - ♦ حاشیه‌نویسی سبکی (Stylistic annotation)
 - ♦ حاشیه‌نویسی واژگانی (Lexical annotation)
- خواهشمند است نام فارسی و انگلیسی حاشیه‌نویسی‌های دیگر را بنویسید.