

**واحد علوم و تحقیقات**

**دانشکده مهندسی مکانیک، برق و کامپیوتر**

**پیشنهاد موضوع رساله دکتری**

**عنوان تحقیق:**

**سیستم پرسش و پاسخ ماژولار با آموزش رقابتی**

استاد راهنما:

سرکار خانم دکتر میترا میرزا رضایی

استاد مشاور:

جناب آقای دکتر وحيد صيدی

**نگارش:**

**مهدی قلیزاده**

**فهرست مطالب**

[**فصل اول : کلیات تحقیق** 6](#_Toc84418708)

[**1-1 مقدمه** 7](#_Toc84418709)

[**2-1 بيان موضوع** 8](#_Toc84418710)

[**3-1 انگیزه تحقيق** 11](#_Toc84418711)

[**4-1 اهمیت و ضرورت انجام تحقیق** 12](#_Toc84418712)

[**5-1 فرضیه­های تحقیق** 12](#_Toc84418714)

[**6-1 سؤالات تحقیق** 13](#_Toc84418715)

[**7-1 جمع بندی** 13](#_Toc84418716)

[**فصل دوم : مروری بر پیشینه پژوهش** 15](#_Toc84418717)

[**1-2- مقدمه** 16](#_Toc84418718)

[**2-2- تاریخچه سیستمهای پرسش و پاسخ** 16](#_Toc84418719)

[**2-3- بازنمایی** 21](#_Toc84418720)

[2-3-1- بازنمایی­های بردار پراکنده 21](#_Toc84418721)

[2-3-2- بازنمایی­های بردار متراکم 25](#_Toc84418722)

[2-3-3- بازنمایی به روش برت 29](#_Toc84418723)

[2-4- پارامترهای ارزیابی 33](#_Toc84418724)

[2-4-1- معیار دقت 34](#_Toc84418725)

[2-4-2- معیار EM 34](#_Toc84418726)

[3-4-2- معیار F1 34](#_Toc84418727)

[4-4-2- معیار ROUGE-L 35](#_Toc84418728)

[5-4-2- معیار BLUE 35](#_Toc84418729)

[**5-2- توابع هزینه مختلف مورد استفاده** 37](#_Toc84418730)

[**6-2- تقسیم­بندی سیستم­های پرسش و پاسخ** 40](#_Toc84418731)

[2-6-1- تقسیم­بندی بر اساس روش طراحی سیستم­های پرسش و پاسخ 40](#_Toc84418732)

[1-1-6-2- سیستم­های مبتنی بر شباهت ظاهری لغات 40](#_Toc84418733)

[2-1-4-2- سیستم­های مبتنی بر نقش کلمات در جمله 41](#_Toc84418734)

[2-3-2- تقسیم­بندی بر اساس نوع منبع اطلاعاتی که پاسخها از آن مشتق شده­اند. 45](#_Toc84418735)

[2-3-3- دسته­بندی سیستم­های پرسش و پاسخ بر اساس حوزه کارکرد 45](#_Toc84418736)

[2-4- معماری سنتی سیستم­های پرسش و پاسخ 48](#_Toc84418737)

[2-4-1- آنالیز پرسش 49](#_Toc84418738)

[2-4-2- بازیابی اسناد 54](#_Toc84418739)

[2-4-3- استخراج پاسخ 56](#_Toc84418740)

[**2-5- کاربرد شبکه‌های عصبی عمیق در سیستمهای پرسش و پاسخ** 57](#_Toc84418741)

[**2-6- سیستم­های پرسش و پاسخ مدرن** 63](#_Toc84418742)

[2-6-1- سیستم­های مبتنی بر خط لوله در سیستمهای پرسش و پاسخ مدرن 64](#_Toc84418743)

[2-6-1-1- بازیاب 64](#_Toc84418744)

[2-6-1-1-1- بازیاب تنک 66](#_Toc84418745)

[2-6-1-1-2- بازیاب متراکم 68](#_Toc84418746)

[2-6-1-1-3- بازیاب تکرارگر 72](#_Toc84418747)

[2-6-1-2- پس پردازش اسناد 75](#_Toc84418748)

[2-6-1-2- خواننده 77](#_Toc84418749)

[2-6-1-2-1- خواننده استخراج­گر 78](#_Toc84418750)

[2-6-1-2-2- خواننده مولّد 80](#_Toc84418751)

[2-6-1-3- پس پردازش پاسخ 81](#_Toc84418752)

[2-6-2- روش‌های پایانه به پایانه 82](#_Toc84418753)

[2-6-2-1- سیستم­های بازیاب – خواننده 83](#_Toc84418754)

[2-6-2-2- سیستم­های بازیاب – تنها 83](#_Toc84418755)

[2-6-2-3- سیستم­های بازیاب – آزاد 84](#_Toc84418756)

[**2-7- علّیت** 85](#_Toc84418757)

[**2-7- چالش‌ها و موانع** 89](#_Toc84418758)

[2-7-1- چالش‌های ساخت سیستم­های پرسش و پاسخ 96](#_Toc84418759)

[2-7-1-1- نظارت از راه دور 96](#_Toc84418760)

[2-7-1-2- بازیابی مؤثر و کارایی 96](#_Toc84418761)

[2-7-1-3- تلفیق دانش 97](#_Toc84418762)

[2-8- سیستم­های پرسش و پاسخ محاوره­ای 99](#_Toc84418763)

[2-9- مجموعه­های داده­ها 100](#_Toc84418764)

[10-2 جمع­بندی 105](#_Toc84418765)

# **فصل اول : کلیات تحقیق**

## **1-1 مقدمه**

با پیشرفت روزافزون تکنولوژی و رخنه ماشین‌های هوشمندی مانند موبایل، کامپیوتر و انواع گجت­ها در زندگی روزمره انسان­ها، نیاز به ارتباط موثر بین ماشین و انسان عمیقا حس می‌شود. انتقال دوطرفه مفاهیم و افکار بین انسان و ماشین، زمینه جذابی است که محققین بسیاری را به خود مجذوب نموده است. هدف از پردازش زبان طبیعی، درک زبان انسان‌ها توسط ماشین‌ها است. اینکه ماشین مانند یک نوزاد زبان را فرا بگیرد و از آن برای ایجاد ارتباط استفاده کند.

آغاز دانش پردازش زبان طبیعی به انتشار مقاله آزمایش تورینگ به قلم آلن تورینگ در سال 1950 میلادی باز می‌گردد. این آزمون به عنوان ملاکی برای سنجش میزان هوشمندی ماشین معروف است. پس از [آزمایش تورینگ](https://fa.wikipedia.org/wiki/%D8%A2%D8%B2%D9%85%D9%88%D9%86_%D8%AA%D9%88%D8%B1%DB%8C%D9%86%DA%AF) فعالیت‌ها بر روی ترجمه زبان‌ها توسط رایانه متمرکز شد. برای مثال در آزمایش جروج‌تاون در سال 1954 میلادی ترجمه خودکار بیش از 60 جمله از زبان روسی به انگلیسی انجام شد.

انتشار کتاب ساختار‌های نحوی اثر نوآم چامسکی زبان‌شناس نام‌آور امریکایی در سال 1957 میلادی، کمک شایانی برای دانشمندان در حوزه پردازش زبان طبیعی بود. تلاش‌های دانشمندان در زمینه پردازش زبان طبیعی از آن زمان تاکنون بی‌وقفه بوده و کاربردهای بسیاری پیدا کرده است.

از جمله کاربردهای بسیار وسیع پردازش زبان طبیعی، پاسخ­گویی به پرسش­های مطرح شده به زبان طبیعی­ است که این زمینه با آغاز دانش پردازش زبان طبیعی که همان آزمون تورینگ است مورد توجه قرار گرفته است.

سیستم‌های پرسش ‎و‎ پاسخ زیر‎شاخه‎ای از علوم پردازش زبان‎طبیعی و بازیابی‎اطلاعات محسوب می‌شوند. x. بنابراین دیگر لازم نیست کاربر مانند سیستم‌های بازیابی‎اطلاعات پرسش خود را به کلید واژه‌‌ها تبدیل کند و پس از بازیابی، در میان تعداد زیادی سند جستجو کند تا به جواب دلخواه خود برسد. در این سیستم­ها به جاي ارائه‎ی کل سند، تنها بخش‎هاي خاصی از اطلاعات سند به عنوان پاسخ بازگردانده می­شوند. این پاسخ ممکن است یک کلمه، یک جمله، یک پاراگراف و یا قطعه­اي صوتی یا تصویري باشد.

سیستم­های پرسش ‎و‎ پاسخ به دهه 1960 بازمی­گردند. سیستم‎های پرسش ‎و ‎پاسخ‎های اولیه بر روی دامنه‎های محدود متمرکز بودند.

دو نمونه قابل توجه از سیستم‎های پرسش ‎و پاسخ، سیستم­های پرسش و پاسخ بیس­بال [3] و لونار [4] می­باشد. سیستم پرسش ‎و پاسخ بیس­بال با هدف پاسخگویی به پرسش­ها درمورد بازي‎هاي بیس­بال انجام شده در لیگ آمریکا در یک فصل، طراحی شده ‎بود. این سیستم یک برنامه‎ی کامپیوتري بود که به پرسش‎هاي مرتبط پاسخ می­داد. پرسش‎ها پیرامون ماه، روز، محل برگزاری، تیم‎ها و امتیازات کسب شده در هریک از بازي­ها از سیستم پرسیده می‎شدند. دومین سیستم لونار بود که آنالیزهای انجام‎ شده روي نمونه سنگ‎های کره ماه که در یک سفر توسط آپولو به دست آمده‎ بودند را در یک پایگاه داده جمع‎آوري کرده و به پرسش‎هایی در این زمینه پاسخ می‎داد. در آن زمان براي ارزیابی این سیستم، زمین‎شناسان را تشویق کردند تا در مدت سه روز، پرسش­های خود را از سیستم بپرسند. از میان 111 سوال پرسیده شده در حوزه‎ي سنگ­هاي ماه، لونار توانست به 78% از پرسش‎ها به درستی پاسخ دهد.

## **2-1 بيان موضوع**

سیستم پرسش و پاسخ یک سیستم خودکار است که جواب پرسش پرسیده شده به زبان طبیعی را با پردازش آن و استفاده از یک پایگاه­داده یا مجموعه­ای از اسناد پیدا می­کند. اهمیت این موضوع در سیستم­های اموزشی یا موتورهای جستجو است که یک کاربر همواره می­خواهد بهترین و مرتبط­ترین جواب ممکن را از مجموعه­ی عظیم داده­ها به دست اورد. سیستم­های پرسش و پاسخ را می­توان از منظرهای مختلفی بررسی نمود که عبارتند از: روش طراحی، نوع منبع اطلاعاتی که پاسخ­ها از آن بدست می­آیند، بر اساس حوزه کارکرد و در نهایت بر اساس نحوه تولید پاسخ خروجی.

از منظر روش طراحی، سیستم­های پرسش و پاسخ به دو دسته مبتنی بر شباهت ظاهری لغات و مبتنی بر نقش کلمات در جمله تقسیم می­شوند. سیستم­های دسته نخست، عمدتاً بر اساس مجموعه کلمات[[1]](#footnote-1) عمل می­کنند. در واقع در نظر گرفتن پرسش به عنوان مجموعه­ای از کلمات، ساده­ترین روش پاسخگویی می­باشد [52]. سیستم­های دسته دوم از ابزارهای مورد استفاده در متن­کاوی و پردازش زبان طبیعی به طور گسترده استفاده می­کنند [1]، [55] و خود به چندین دسته تقسیم می­شوند که در فصل آینده مفصلاً به آنها پرداخته شده است.

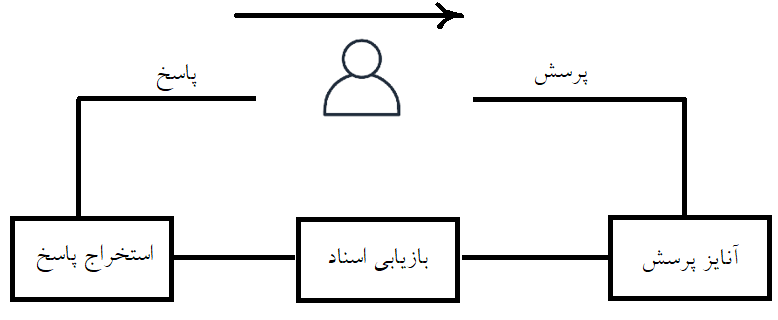
از منظر نوع منبع اطلاعاتی که پاسخ­ها از آن بدست می­آیند، سیستم­های پرسش و پاسخ به دو دسته مبتنی بر اسناد و مبتنی بر پایگاه دانش تقسیم­بندی می­شوند. سیستم­های پرسش و پاسخ مبتنی بر اسناد، از اسناد متنی بدون ساختار برای پاسخگویی استفاده می­کند، در حالی که سیستم­های پرسش و پاسخ مبتنی بر پایگاه دانش[[2]](#footnote-2) از پایگاه­های دانش ساختاریافته و از پیش تعریف شده مانند انواع پایگاه­داده و پایگاه دانش که اغلب به صورت دستی ساخته می­شوند، برای پاسخگویی استفاده می­کنند. سیستم­های پرسش و پاسخ متنی در مقایسه با سیستم­های پرسش و پاسخ مبتنی بر پایگاه دانش بسیار مقیاس­پذیرتر هستند، زیرا منابع متنی بدون ساختار نسبتاً رایج بوده و به راحتی در دسترس هستند. مانند: ویکی­پدیا، اخبار، مقالات، کتب علمی و غیره. در حالی که پایگاه­های دانش، نیاز به صرف زمان و منابع مختلف زیادی برای تولید دارند.

از منظر حوزه کارکرد سیستم­های پرسش و پاسخ به دو دسته سیستم­های پرسش و پاسخ با دامنه محدود[[3]](#footnote-3) و سیستم­های پرسش و پاسخ با دامنه باز[[4]](#footnote-4) تقسیم می­شوند. این دو سیستم از جهات گوناگونی با هم متفاوتند اما به طور کلی سیستم­های دامنه محدود [64] تنها در یک حوزه معین و مشخص توانایی پاسخگویی دارند. برای مثال سیستم­هایی که توانایی پاسخگویی در مسائل پزشکی را دارند و در دیگر حوزه­ها ناتوانند. اما سیستم­های دامنه باز [54] محدود به حوزه خاصی نیستند و توانایی پاسخگویی به پرسش­ها را به صورت عمومی دارند.

از منظر نحوه تولید پاسخ، سیستم­های پرسش به دو نوع سیستم­های استخراج­گر و سیستم­های مولّد تقسیم می­شوند. پیش فرض سیستم­های استخراج­گر این است که پاسخ صحیح به پرسش ورودی قطعاً در اسناد وجود دارد، و معمولاً بر یادگیری برای پیش‌بینی موقعیت شروع و پایان یک پاسخ از اسناد بازیابی شده تمرکز دارند. روش­های موجود در این حوزه با توجه به اینکه آیا اسناد بازیابی شده به طور مستقل پردازش می‌شوند یا به طور مشترک، تقسیم­بندی می­شوند که در فصل 2 به آنها پرداخته شده است [‏29]‏، [47]‏، [142]. سیستم­های مولّد به جای استخراج محدوده پاسخ که معمولاً بر مدل‌های دنباله به دنباله متکی هستند، تا حد امکان پاسخ‌های طبیعی، تولید می­کنند [159]، [162].

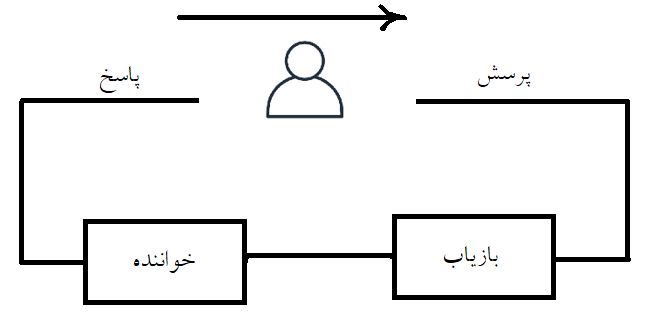
معماری­های مختلفی برای سیستم­های پرسش و پاسخ ارائه شده است. دو معماری اصلی سیستم­های پرسش و پاسخ عبارتند از: معماری سنتی و معماری مدرن.

معماری سنتی در شکل 1-1 قابل مشاهده است. این معماری سه بخش اصلی آنالیز پرسش، بازیابی اسناد و استخراج پاسخ تشکیل شده است.



شکل 1-1- نمونه­ای از معماری سنتی در سیستم­های پرسش و پاسخ

و معماری مدرن که از دو بخش بازیاب و خواننده تشکیل شده است و در شکل 2-2 قابل مشاهده می­باشد.



شکل 1-2- نمونه­ای از معماری "بازیاب – خواننده" در سیستم­های پرسش و پاسخ

سیستم­های پرسش و پاسخ فراوانی با استفاده از این معماری­ها و انواع متدهای یادگیری عمیق مانند [[5]](#footnote-5)LSTM، [[6]](#footnote-6)GRU، انواع رمزگذارها[[7]](#footnote-7) و ... پیشنهاد شده­اند که در قسمت پیشینه تحقیق به آنها پرداخته شده است.

سیستم­های پرسش و پاسخ به طور معمول سعی می­کنند با استفاده از یک مدل قدرتمند به خوبی آموزش دیده پاسخ­ها را تولید نمایند. از منظر این مدل­های قدرتمند، ماهیت تمامی پرسش­ها یکی بوده و هیچ تفاوتی بین آنها وجود ندارد و یا در بهترین حالت، در معماری سنتی سیستم­های پرسش و پاسخ در بخش آنالیز پرسش تنها تفاوت بین پرسش­ها را نوع پاسخ مورد انتظار آنها، در نظر می­گیرند. حال آنکه پرسش­ها می­توانند از منظر­های گوناگونی مانند معنا، حوزه پرسش، زمان پرسش، نحوه پرسیدن پرسش، نوع پاسخ و... متفاوت باشند. چنانچه این موارد در یک سیستم پرسش و پاسخ در نظر گرفته شود، پاسخ­های دقیق­تر و مناسب­تری تولید خواهد شد. اما از آنجایی که پرسش­ها می­توانند از جنبه­های بسیار زیادی متفاوت باشند، لذا شناسایی جنبه­ها و دسته­بندی پرسش­ها عملا غیر ممکن است. لذا نیاز به روشی است که به صورت خودکار زمینه پرسش ورودی را تشخیص داده و پاسخ مناسب را تولید نماید. برای این منظور، چنانچه بتوان به جای استفاده از یک مدل قدرتمند، از چند مدل کوچکتر به صورت موازی استفاده نمود که هریک از مدل­­ها برای پاسخگویی به حوزه خاصی متخصص گردد، آنگاه می­توان به پاسخ­های مناسب تری دست یافت. با توجه به مطالب فوق­الذکر هدف از این تحقیق پیشنهاد یک سیستم پرسش و پاسخ ماژولار با آموزش رقابتی است. در این سیستم هر ماژول یک متخصص است. متخصصین برای کسب تخصص برای تولید پاسخ حوزه خاصی به رقابت با یکدیگر می­پردازند. این امر سبب می­گردد تا به جای یک مدل از چند مدل استفاده شود و همچنین رقابت بین متخصص­ها باری تولید پاسخ باعث متخصص شدن آنها در زمینه­­های مختلف خواهد شد.

## **3-1 انگیزه تحقيق**

تفکّر غالب بر سیستم­های پرسش و پاسخ، استفاده از یک "کلید طلایی" برای حل تمامی مسائل می­باشد. منظور از کلید طلایی استفاده از یک راه حل جامع برای پاسخگویی به تمامی پرسش­ها است. لذا مهمترین مانعی که در سیستم­های پرسش و پاسخ مطرح شده، قابل مشاهده است، عدم ثبات در کارآیی در مجموعه داده­های مختلف می­باشد. ممکن است روشی پیشنهاد شود که در یک مجموعه داده کارآیی مناسبی داشته باشد. اما در مجموعه­های دیگر کارآیی مناسبی نداشته و می­بایست مجدّدا در اکثر موارد از صفر آموزش داده شوند. البته روش­هایی مانند انتقال یادگیری مطرح شده است که فرآیند آموزش برای مجموعه داده جدید را تسریع می­نماید. از سوی دیگر در نظر گرفتن جنبه­های مختلف یک پرسش می­تواند کمک بسزایی در درک مفهوم پرسش، و تولید پاسخ مناسب­تر نماید. اما از آنجایی که پرسش­ها می­توانند از جنبه­های بسیار زیادی متفاوت باشند، لذا شناسایی جنبه­ها و دسته­بندی پرسش­ها بر اساس آنها عملا غیر ممکن است. انگیزه این تحقیق ارائه روشی برای حل مشکلات مطرح شده است که به جای یک سیستم پرسش و پاسخ خیلی قوی با هزینه آموزش بالا از چند مدل کوچکتر و با هزینه آموزش کمتر به صورت موازی استفاده شود.

## **4-1 اهمیت و ضرورت انجام تحقیق**

سیستم­های پرسش و پاسخ، کاربردهای وسیعی برای افراد، شرکت­های تجاری و تبلیغاتی و امور نظامی دارند. این سیستم­ها با ارائه پاسخ دقیق و تا حد امکان کوتاه زمان یافتن پاسخ­ها را کوتاه می­نمایند. مدل­های پرسش و پاسخ عموماً سعی دارند با استفاده از یک مدل قوی به پرسش­های مطرح شده پاسخ دهند. چنین مدل­هایی نه تنها پارامترها و هایپر پارامترهای زیاد و زمان آموزش بالایی دارند، بلکه پاسخ­های تولید شده­ی آنها نیز چندان رضایت بخش نمی­باشند.

در حقیقت گستردگی و تنوع بسیار بالا در پرسش­های مطرح شده و پاسخ مورد انتظار، ایجاب می­کند که یا همچنان از مدل­­های پیچیده­تر با تعداد پارامترهای زیاد برای تولید پاسخ مناسب­تر استفاده شود (که این موضوع هم نیاز به سخت افزار و امکانات بیشتری دارد و هم زمان آموزش را به صورت نمایی افزایش می­دهد). و یا از روش هوشمندتری برای تولید پاسخ­ها استفاده شود. در این روش از چندین مدل برای تولید پاسخ­ها استفاده می­شود که هر یک از مدل­ها در زمینه خاصی متخصص شده و مسئول پاسخگویی به پرسش­های مطرح شده در حوزه تخصصی خود می­باشد. از آنجایی که این مدل­های متخصص، قابلیت آموزش موازی دارند، لذا به جای آموزش یک مدل بزرگ، از چند مدل کوچکتر که به صورت موازی آموزش می­بینند، استفاده می­شود، که از یک سو باعث کوتاه شدن زمان آموزش و از سوی دیگر به دلیل متخصص شدن مدل­های کوچک در حوزه­های گوناگون، باعث دقیق­تر شدن پاسخ­ها می­گردد. بنابراین با توجه به موارد فوق­الذکر دو فاکتور زمان آموزش کوتاه­تر و تولید پاسخ دقیق­تر این روش در قیاس با روش­های موجود، ضرورت انجام آن را توجیه می­نماید.

## **5-1 فرضیه­های تحقیق**

* استفاده از روش ترکیب مدل­های پرسش و پاسخ منجر به تولید پاسخ­­های دقیق­تر و کارآمدتری خواهد شد.
* آموزش موازی در مدل ترکیبی سیستم­های پرسش و پاسخ فرآیند آموزش را کاهش می­دهد.
* در سیستم­های پرسش و پاسخ، می­توان با بهره­گیری از ترکیب چند مدل ساده­تر و یک فرآیند آموزش رقابتی به پاسخ­های بهتری دست یافت.

## **6-1 سؤالات تحقیق**

* استفاده از ترکیب چند مدل ساده به جای یک مدل بزرگ چه تاثیری در کارآمدی سیستم­های پرسش و پاسخ دارد؟
* استفاده از آموزش رقابتی مدل­ها در سیستم­های پرسش و پاسخ چه تاثیر در بهینه­تر شدن پاسخ­ها خواهد داشت؟

## **7-1 اهداف مشخص تحقيق (شامل اهداف آرماني، کلی، اهداف ويژه و كاربردي):**

سیستم­های پرسش و پاسخ در زمینه­های بسیاری دارای کاربردهای گوناگونی است که در هر یک از آنها اهداف متفاوتی دنبال می­شود. در اینجا به ذکر اهداف مهمی که این مستند به دنبال آن است پرداخته می­شود.

هدف آرمانی این مستند، ارائه یک سیستم پرسش و پاسخ که قادر به تولید پاسخ­های دقیق و تا حد امکان کوتاه برای پرسش­هایی که مطرح شده است، می­باشد. اما به صورت ویژه این مستند به دنبال اهداف ذیل است.

* با استفاده از ترکیب چند مدل پرسش و پاسخ و آموزش موازی آنها به جای یک مدل بزرگ زمان آموزش را کاهش دهد.
* آموزش مدل­ها، به صورت رقابتی انجام شود که منجر به متخصص شدن مدل­ها در حوزه های به خصوص و در نتیجه ارائه پاسخ­های دقیق­تر ومرتبط­تر شود.
* یک سیستم ماژولار پایانه به پایانه ارائه نماید که در آن اولا، ماژول­ها قابلیت استفاده مجدد داشته باشند و دوما به صورت پایانه به پایانه آموزش ببینند.

## **8-1 نوآوری تحقیق**

سیستم­های پرسش پاسخ دارای کاربردهای فراوانی در حوزه­های مختلف از جمله نظامی، آموزشی، تحقیقاتی و ... می­باشند. استفاده از یک مدل بزرگ و پیچیده برای پاسخگویی به پرسش­ها منجر به نتایج نه چندان رضایت بخشی خواهد شد. بنابراین در این مطالعه سه معماری جدید برای سیستم­های پرسش و پاسخ پیشنهاد می­شود که هر سه آنها بر پایه ترکیب مدل­های ساده­تر و آموزش رقابتی آنها می­باشد. پیشنهاد نخست، استفاده از مدل­ها متفاوت و مستقل پرسش و پاسخ به گونه­ای که در یک فرآیند آموزشی رقابتی برای کسب تخصص آموزش ببینند. در پیشنهاد دوم، از مدل­های متفاوت پرسش و پاسخ استفاده می­شود. این مدل­ها مستقل نبوده و با یکدیگر تبادل اطلاعات دارند. یعنی با به اشتراک گذاشتن دانش خود با دیگر متخصصین که از طریق اشتراک گذاری وزن­ها صورت می­گیرد در فرآیند آموزشی رقابتی برای کسب تخصص آموزش می­بینند. نهایتا در پیشنهاد سوم، متخصصین در یک ساختار به نام رمزگذار خودکار متغیر چیده می­شوند. در این روش بر خلاف دو روش پیشین به تعداد متخصصین، مدل وجود ندارد و تنها یک مدل رمزگذار خودکار متغیر وجود دارد که متخصصین در آن چیده می­شوند. در واقع این روش هر یک از توزیع­های به دست آمده در لایه پنهان رمزگذار خودکار متغیر را به عنوان یک متخصص تولید پاسخ در نظر می­گیرد.

## **9-1 ساختار مستند**

ساختار این پروپوزال به این صورت دسته­بندی شده است؛ در فصل دوم مروری بر مقالات انجام شده در حوزه ­های مختلف سیستم­های پرسش و پاسخ خواهد شد و یک دسته­بندی از روش­های به کار گرفته شده صورت خواهد گرفت، در انتهای این فصل در مورد مجموعه داده­های مورد استفاده بحث می­شود. در فصل سوم هم روش پیشنهادی پروپوزال پیش رو مطرح خواهد شد.

# **فصل دوم : مروری بر پیشینه پژوهش**

## **1-2- مقدمه**

پاسخگویی به پرسش­­ یکی از شاخه­های بسیار مهم و کاربردی در حوزه پردازش زبان طبیعی به شمار می­آید و هدف اصلی آن، ارائه پاسخ مناسب به پرسش­های­ مطرح شده به زبان طبیعی از اسناد بدون ساختار می­باشد. در مقایسه با یک موتور جستجو، هدف سیستم پرسش و پاسخ، ارائه پاسخ مشخص و دقیق به یک پرسش، به جای بازگشت لیستی از لینک­ها یا ابرلینک­هاست. در نتیجه می­تواند برای کاربر خیلی مؤثر و دلنشین باشد. امروزه بسیاری از موتورهای جستجوی وب مانند گوگل[[8]](#footnote-8) و بینگ[[9]](#footnote-9) در حال تکامل به ارائه پاسخ­های دقیق به پرسش‌های مورد نظر کاربر هستند. اخیراً افزایش قابل توجهی در میزان مقالات تحقیقاتی در سیستم­های پرسش و پاسخ، به ویژه در مورد تکنیک­هایی که با درک مطلب ماشین[[10]](#footnote-10) ادغام شده­اند، مشاهده می‌شود. با استفاده از این تکنیک­ها، موتورهای جستجو قادر به پاسخگویی دقیق به برخی از انواع پرسش­ها از قبیل پرسشی که در ذیل آمده است، می­باشد:

پرسش: چه زمانی انقلاب اسلامی در ایران پیروز شد؟ پاسخ: 22 بهمن ۱۳۵۷

این فصل سیستم­های پرسش و پاسخ را با جزئیات بیشتری مورد بررسی قرار می­دهد. در ابتدا تاریخچه­ای از سیستم­های پرسش و پاسخ بیان می­شود و در آن، این سیستم­ها از ابتدایی­ترین حالت تا سیستم­های پرسش و پاسخ هوشمند کنونی به طور اجمالی معرفی می­شود. در ادامه به بحث پیرامون انواع بازنمایی­های استفاده شده، انواع معیارهای مناسب برای ارزیابی کارایی و توابع هزینه مورد استفاده برای آموزش سیستم­های پرسش و پاسخ پرداخته می­شود. سپس سیستم­های پرسش و پاسخ سنتی بررسی می­شود و در ادامه تاثیر یادگیری عمیق در بهبود سیستم­های پرسش و پاسخ و در نهایت مدل درک مطلب ماشین عصبی مورد تحلیل و آنالیز قرار می­گیرد.

## **2-2- تاریخچه سیستم­های پرسش و پاسخ**

طراحی سیستم­های پرسش و پاسخ به عنوان رابطی برای دسترسی به داده­ها یا پایگاه­های دانش، یکی از زمینه­های تحقیقاتی مهم در حوزه پردازش زبان طبیعی می­باشد. تحقیقات برای طراحی سیستم­هایی که قادر به پاسخگویی به پرسش­های کاربر باشند از سال 1950، توسط آلن تورینگ آغاز شد. آلن تورینگ تفکر پرسش و پاسخ را به کمک "بازی تقلید[[11]](#footnote-11)" که با نام تست تورینگ[[12]](#footnote-12) شناخته می­شود، معرفی کرد. در این تست انسان از کامپیوتر می­پرسد و کامپیوتر پاسخ می­گوید، که این زمینه­ساز نگرشی نو، برای ایجاد سیستم­های پرسش و پاسخ هوشمند گردید.

یک سیستم­ پرسش و پاسخ، موتور جستجویی است که پرسش­های به زبان طبیعی را به عنوان ورودی دریافت کرده و پاسخ متناسب و دقیق را به کاربر ارائه می­نماید. سیستم­های پرسش و پاسخ بر اساس دامنه عملکردشان به دو حوزه دامنه باز و دامنه محدود دسته­بندی می­شوند. شمای کلی یک سیستم پرسش و پاسخ در شکل 2-1 قابل مشاهده است.

|  |  |  |
| --- | --- | --- |
| پرسش: امیرکبیر کی متولد شد؟ |  | ۲۲ آبان ۱۲۳۰ خورشیدی |

شکل 2-1- تصویری از سیستم پرسش و پاسخ. این سیستم با توجه به پرسش زبان طبیعی، پاسخ را از مجموعه­ای از اسناد متنی بدون ساختار، استنتاج می­کند.

در اواخر دهه1950 میلادی اولین سیستم­های واقعی پرسش و پاسخ معرفی شدند که عمدتاً در حوزه دامنه محدود و مبتنی بر الگوی کلمات موجود در ورودی بودند. به عنوان نمونه از این سیستم­ها، می­توان به ماشین محاوره[[13]](#footnote-13) معرفی شده توسط آقای گرین[[14]](#footnote-14) [2] اشاره کرد. از آنجایی که این سیستم­ها بسیار محدود بودند کاربران خیلی سریع متوجه می­شدند که با یک ماشین سرو کار دارند، نه یک انسان واقعی.

در اوایل دهه 1960 میلادی توجهات به سمت استفاده از پایگاه­های داده­ برای تولید سیستم­های پرسش و پاسخ به عنوان واسط زبان طبیعی، جلب شد. دلیل این عنوان این است که منبع اطلاعاتی این سیستم­ها، یک پایگاه داده از اطلاعات مرتبط با حوزه کاری مد­نظر آنها می­باشد. هدف این سیستم­ها تبدیل هر چه دقیق­تر پرسش کاربر به پرس­وجوی پایگاه داده و ارائه نتیجه پرس­وجو به عنوان پاسخ به کاربر است. بیس­بال[[15]](#footnote-15) [3] و لونار[[16]](#footnote-16) [4] از جمله چنین سیستم­هایی هستند. سیستم پرسش و پاسخ بیس بال توسط آقای گرین [5] معرفی شد. این سیستم طراحی شد تا در مورد جزئیات بازی­های بیس­بال آمریکا مانند نتایج، مکان و تاریخ بازی­ها در طول یک فصل پاسخ بگوید. در این سیستم از تکنیک تجزیه سطحی[[17]](#footnote-17) برای شناسایی موجودیت­ها، مانند: نام تیم­ها و آمار موجود در پرسش­ها استفاده می­شد. سیستم لونار توسط آقای وودز[[18]](#footnote-18) معرفی شد. پایگاه داده استفاده شده در این سیستم دارای اطلاعاتی راجع به صخره­های کره ماه بود و بنابراین قادر به پاسخگویی پرسش­های مطرح شده توسط متخصصین زمین­شناسی تا حدود 90 درصد شد. با وجود اینکه سیستم­های مذکور عملکرد مناسبی به پرسش­ها در محدوده خود داشتند، اما از آنجایی که پاسخ­ها کاملا محدود به پایگاه داده و ساختار مورد استفاده آن بود لذا، در خصوص پرسش­هایی که سیستم برای پاسخگویی به آنها طراحی نشده بود، بسیار ضعیف عمل می­کردند.

شخصی به نام وینگوراد[[19]](#footnote-19) سیستم پرسش و پاسخی به نام شردلو[[20]](#footnote-20) را درسال 1972 ارائه نمود [6]. این سیستم علی­رغم اینکه کارایی مناسبی در درک زبان طبیعی از خود نشان داد، اما در تست تورینگ عملکرد ضعیفی داشت. در اواسط دهه 1970 میلادی سیستم مارگی[[21]](#footnote-21) توسط چنک[[22]](#footnote-22) [7] ارائه شد که می­توانست اسناد را بخواند، تفسیر کند و به پرسش­هایی در مورد آن اسناد پاسخ بگوید. سیستم مارگی بر اساس ایده نحوه سازماندهی اطلاعات در مغز انسان ساخته شد و درک مطلب را به وسیله تجزیه متن به بازنمایی معنایی انجام می­داد. به طور کلی تلاش­های صورت گرفته در دهه 1970 میلادی، نشان از شروع حرکت­هایی برای درک مکالمات انسانی است.

سیستم­های مبتنی بر دانش با دامنه محدود، در دهه­های 1980 و 1990 غالب توجهات محققین را به خود معطوف کرد. این فن­آوری به طور گسترده­ای در سیستم­های پرسش و پاسخ که در آنها کاربران با نیازهای خاصی مواجه بودند، مورد استفاده قرار گرفت. در این روش، از دانش موجود در پایگاه دانش و یک سری اطلاعات اضافی ایجاد شده توسط کاربر استفاده ­شد و دسترسی به پایگاه دانش از طریق یک رابط زبان طبیعی انجام می­شد. همچنین با هدف درک بهتر مقصود کاربر، پرسش­های اضافی نیز از کاربر پرسیده می­شد. مایسین[[23]](#footnote-23) [8] یک سیستم پرسش و پاسخ مبتنی بر دانش با دامنه محدود است که برای توضیحات مفاهیم پزشکی طراحی شده بود. هنوز نیز در پرسش­های به زبان طبیعی توجه ویژه­ای به پرس­وجو از پایگاه دانش می­شود [9].

سال­های واپسین دهه 1990 را می توان، عصر جدید سیستم­های پرسش و پاسخ دانست که از منظر روش طراحی می­توان آنها را به دو دسته سیستم­های مبتنی بر شباهت ظاهری لغات و سیستم­های مبتنی بر نقش کلمات در جمله تقسیم­بندی کرد.

در سال ۱۹۹۳ موراکس[[24]](#footnote-24) [10] برای پاسخ به پرسش‌های ساده دانشگاهی بر اساس یک دانشنامه دانشگاهی انگلیسی طراحی شد که عمدتاً از تحلیل زبانی و تکنیک­های تطبیق الگوی نحوی استفاده می­کرد. برای نخستین بار، در سال ۱۹۹۹ در کنفرانس بازیابی اسناد، یک سیستم پرسش و پاسخ دامنه باز، به منظور استخراج ۵ بخش برتر، از مجموعه مقالات خبری که احتمالاً حاوی پاسخ صحیح بودند، معرفی شد (TREC) [11]. در سال­های بعد مجموعه­ای از تلاش­ها، تحت عنوان TREC QA باعث پیشرفت چشمگیری در حوزه سیستم­های پرسش و پاسخ متن­باز شدند [12]، [13] و [14].

کمپین TREC مجموعه محلی از اسناد را به عنوان یک منبع اطلاعاتی جهت تولید پاسخ در اختیار محققین قرار می‌داد. اما به دلیل محبوبیت شبکه جهانی وب و به ویژه افزایش کارآمدی و بلوغ موتورهای جستجو، محققان به ساخت سیستم­های پرسش و پاسخ متن باز مبتنی بر وب، با جمع­آوری پاسخ­ها از منابع آنلاین نظیر Google.com و Ask.com و با استفاده از تکنیک­های بازیابی اطلاعات، روی آوردند [15]، [16]، [17]، [18]. موتورهای جستجوی وب توانایی جمع‌آوری مداوم و مؤثر صفحات وب را به صورت انبوه دارا هستند. از این رو قادر به ارائه اطلاعات بیشتر و مؤثرتر برای کمک به پاسخگویی به پرسش­های کاربر می­باشند. در سال ۲۰۰۱، یک سیستم پرسش و پاسخ به نام مولدر[[25]](#footnote-25) [15] برای پاسخ خودکار به پرسش‌های از نوع حقیقی[[26]](#footnote-26) در دامنه باز به کمک یک موتور جستجو (برای مثال google.com) طراحی شد. در این روش ابتدا پرسش­های کاربران با استفاده از تعدادی تجزیه­گر زبان طبیعی به چندین پرس­وجو ترجمه می‌شد. سپس پرس­وجوهای بدست آمده به موتور جستجو ارسال می­شد تا از طریق آنها اسناد مرتبط استخراج شوند. سپس از بخش استخراج پاسخ، برای استخراج پاسخ از نتایج بازگشتی استفاده می­شد.

در آغاز قرن 21 میلادی اکثر تحقیقات در سیستم­های پرسش و پاسخ با مفاهیمی مانند پردازش زبان طبیعی، بازیابی اطلاعات و استخراج اطلاعات سر و کار دارد [19]، [20]، [21]، [22]. سیستم­های پرسش و پاسخ سنّتی عمدتاً از یک خط لوله[[27]](#footnote-27) متشکل از سه مرحله، تحلیل پرسش، بازیابی سند و استخراج پاسخ پیروی می­کنند [23]، [21]، [24]. از آنجایی که پرسش ورودی به زبان طبیعی مطرح می­شود، لذا هدف از تجزیه و تحلیل پرسش این است که پرسش مجدداً تدوین شود تا پرس­وجو­های بهتر و دقیق­تری برای بازیابی اسناد در مرحله بعد مهیّا گردد. همچنین در این بخش، طبقه‌بندی پرسش انجام می­شود تا نوع پاسخ مورد انتظار پرسش استخراج شود.  در مرحله بازیابی اسناد، سیستم به جستجوی اسناد یا متون مرتبط با پرسش با استفاده از تکنیک­های موجود مانند TF-IDF وBM25، یا تکنیک­های خاص توسعه­یافته برای موتورهای جستجوی وب مانند google.com و یا Bing.com می­پردازد. سپس در مرحله استخراج پاسخ، جواب نهایی از اسناد و اطلاعات به دست آمده از مرحله قبل استخراج می­شوند.

تکنیک­های یادگیری عمیق، که پیشرفت قابل توجهی در بسیاری از زمینه‌ها داشته­اند، تقریباً در هر مرحله از سیستم­های پرسش و پاسخ قابلیت به کارگرفته شدن را دارند [25]. به عنوان مثال لئی[[28]](#footnote-28) و همکاران در سال 2018 [26] و همچنین ژیا[[29]](#footnote-29) و همکاران در سال 2018 [27] ، طبقه­بندی کننده پرسش را با استفاده از یک مدل مبتنی بر CNN[[30]](#footnote-30) و یک مدل مبتنی بر LSTM[[31]](#footnote-31) توسعه می­دهند. همچنین سیستم­های ارائه شده توسط نیشیدا در سال 2018 [28]، کارپوخین در سال 2020 [29] و خطیب در سال 2020 [30] که همگی به تفصیل در بخش 2-6-1-1-2 برسی شده­اند، چند مدل بازیابی اطلاعات، با استفاده از شبکه­های عصبی عمیق برای جستجوی اسناد مرتبط در یک فضای نهفته[[32]](#footnote-32) را پیشنهاد می­کنند. در سال­های اخیر با ظهور برخی از مجموعه داده­های مقیاس بزرگ مرتبط با سیستم­های پرسش و پاسخ [31]، [32]، [33]، [34]، [35]، [36]، تکنیک‌های درک مطلب ماشین عصبی تا حد زیادی پیشرفت کرده­اند [31]، [37]، [38]، [39] و [40] و انقلابی در سیستم­های پرسش و پاسخ سنتی به وجود آورده­اند [41]، [42]، [43]، [44] که منجر به ظهور یک معماری مدرن " بازیاب – خواننده[[33]](#footnote-33)" شده است. بازیاب، به عنوان سیستم بازیابی اطلاعات، مسئول بازیابی اسناد مرتبط با پرسش ورودی است و این در حالی است که هدف از بخش خواننده، که معمولاً به صورت یک مدل درک مطلب ماشین عصبی است، استنتاج جواب نهایی از اسناد بازیابی شده می­باشد. در بعضی مقالات حتی نام سیستم پرسش و پاسخ به ماشین درک مطلب ماشین مقیاس پذیر تغییر یافته است [41] و [45] و [46]. به دنبال این معماری، تحقیقات گسترده­ای در جهات مختلف از قبیل: رتبه­بندی مجدد اسناد بازیابی شده [42]، [47]، [48]، بازیابی اسناد مرتبط با یک پرسش به صورت تکرارگر [29]، [35]، [36] و آموزش کل سیستم پرسش و پاسخ به شیوه پایانه به پایانه[[34]](#footnote-34) [28]، [44]، [49]، [50] و غیره صورت گرفته است.

## **2-3- بازنمایی**

بدیهی است که کامپیوترها تنها قادر به درک زبان باینری، به صورت 0 و 1 می­باشند و توانایی درک کلمات زبان طبیعی را ندارند. اما کدگذاری کلمات به شکل عددی می­تواند مشکل را حل نماید.

فرآیند تبدیل اطلاعات متنی به اعداد را وکتورایز کردن[[35]](#footnote-35) می­نامند. روش­های مختلفی برای تبدیل متن به اعداد وجود دارند که عبارتند از: نمایش­های بردار پراکنده[[36]](#footnote-36) و نمایش­های بردار متراکم[[37]](#footnote-37).

### 2-3-1- بازنمایی­های بردار پراکنده

**الف) بازنمایی به روش کیف کلمات**[[38]](#footnote-38)

در روش بازنمایی به روش کیف کلمات [123]، کلمات منحصر به فرد هر متن، کیف کلمات متن مورد نظر را تشکیل می­دهند. نمونه­ای از کیف کلمات در مثال زیر قابل مشاهد است.

|  |  |  |
| --- | --- | --- |
| متن: من عاشق عسل هستم. من عاشق پول هستم. |  | کیف کلمات: من، عاشق، عسل، پول |

کیف کلمات یک الگوریتم است که تعداد دفعات نمایش یک کلمه در یک سند را محاسبه می­کند. این تعداد کلمات محاسبه شده، امکان مقایسه اسناد و شباهت آنها را برای برنامه­هایی مانند: جستجو، طبقه­بندی اسناد و مدل­سازی موضوع فراهم می­نماید. این نام به این دلیل انتخاب شده که فقط مربوط به وقوع کلمه است و نه جایی که در کیف قرار می گیرد. یعنی در این روش اصلاً ترتیب قرارگیری کلمات اهمیّتی ندارد. توجیه چنین رویکردی این است که اسناد مشابه حاوی کلمات مشابه هستند.

در رویکرد فوق، هر کلمه یا نشانه "گرام[[39]](#footnote-39)" نامیده می­شود. یک روش بهتر تولید کیف کلمات، ایجاد واژگان کلمات گروه­بندی شده است که دامنه واژگان را تغییر داده و در نتیجه مجموعه کلمات، معنای بیشتری از اسناد را اخذ می­کند. ایجاد واژگان از جفت­های دو کلمه­ای را مدل بایگرام و واژگان سه­گانه کلمات را مدل سه­گرام می­نامند. رویکرد کلی مدل n-gram نامیده می­شود، که در آن n به تعداد کلمات گروه­بندی شده اشاره می­کند. مزیت روش کیف کلمات:

* فهم ساده و پیاده­سازی آسان مزیت این روش است.

معایب روش کیف کلمات:

* بازنمایی بزرگ و اطلاعات کم: از آنجایی که واژگان، شامل همه کلمات منحصر به فرد موجود در متون داده شده است؛ لذا در متون بزرگ حاوی تعداد زیادی کلمات منحصر به فرد، باعث تولید بازنمایی­های با فضای بزرگ و اطلاعات کم خواهد شد.
* عدم وجود ترتیب در کلمات: مدل کیف کلمات فقط مربوط به وقوع کلمه است و نه جایی که در کیف قرار داده شده است (یعنی ترتیب). این امر منجر به از دست رفتن اطلاعات زمینه­ای و در نتیجه معنای کلمات موجود در سند (معناشناسی) می شود. به عنوان مثال مدل در درک جملاتی که کلمات یکسانی دارند ولی ترتیب قرارگیری کلمات متفاوت است، عاجز می­باشد. همچنین توانایی در درک مترادف­ها وجود ندارد. برای مثال دوچرخه قدیمی و دوچرخه کهنه از هم متفاوت تشخیص داده می­شوند.

**ب) بازنمایی به روشTF-IDF**

روش بازنمایی TF-IDF [126]، بر مبنای دو مفهوم تعداد اصطلاحات[[40]](#footnote-40) و معکوس تعداد اسناد[[41]](#footnote-41) بنا نهاده شده است.

در یک سند مشخص، اهمیت یک کلمه با توجه به تعداد دفعاتی که در سند ظاهر می­شود، تعیین می­گردد. تعداد اصطلاحات، اهمیت کلمه را تعیین می­کند. اگر یک کلمه بارها ظاهر می­شود، آن کلمه باید مهم باشد.

معکوس تعداد اسناد، برای محاسبه وزن کلمات کمیاب در همه اسناد استفاده می­شود. کلماتی که به ندرت در متون یافت می­شوند دارای امتیاز IDF بالا هستند. با این حال، مشخص است که عبارات خاصی ، مانند "من"، "و" ممکن است بارها ظاهر شوند ، اما اهمیت کمی دارند. بنابراین نیاز است وزن اصطلاحات پر تکرار کم و در عین حال وزن اصطلاحات نادر افزایش یابد.

برای به دست آوردن ضریب TF-IDF می­بایست هر کدام از دو عبارت فوق را به صورت جداگانه محاسبه نموده و حاصل دو عبارت را در هم ضرب کنیم تا نتیجه حاصله، فراوانی وزنی کلمه کلیدی را به ما نشان دهد.

فرمول دو عبارت فوق به شرح زیر می باشد:

TF: عبارت است از تقسیم تعداد تکرار کلمه بر تعداد کل کلمات محتوا.

IDF: عبارت است از لگاریتم تقسیم تعداد کل محتوا بر محتواهایی که شامل کلمه مورد نظر هستند.

این بازنمایی در رابطه 2-1 قابل مشاهده است.

|  |  |
| --- | --- |
|  | رابطه 2-1 |

: تعداد تکرار x در y.

: تعداد اسنادی که حاوی x هستند.

N: تعداد کل اسناد.

**ج) بازنمایی به روش BM25**

روش BM25 [137]، یک روش­ رتبه­بندی مبتنی بر محتوا در مدل­های احتمالی می­باشد. هدف از مدل ارتباط احتمالی، پیدا کردن احتمال وابستگی هر سند به هر پرس­وجو می­باشد. به عبارت دیگر، هدف، یافتن احتمال مرتبط بودن هر سند با پرس­وجوی کاربر است. بنابراین قادر به یافتن درجه شباهت میان پرس­وجو و سند می­باشد. هسته اصلی این مدل استفاده از تئوری احتمال است.

آقای رابرتسون[[42]](#footnote-42) یک مکانیزم وزن­گذاری مبتنی بر 2-پواسون به نام اُکاپی[[43]](#footnote-43) که دارای ویرایش­های متنوعی به صورت BMnn است را ارائه کرده است. نامگذاری این مکانیزم (2-پواسون) به خاطر این است که توزیع هر واژه در یک مجموعه اسناد دارای توزیع پواسون می­باشد (بدیهی است که تکرار هر واژه در هر مجموعه­ی اسناد یک توزیع دوجمله­ای است و در شرایطی که n[[44]](#footnote-44) زیاد و P[[45]](#footnote-45) کم باشد، توزیع دوجمله­ای به پواسون تبدیل می­شود). این روش احتمالی از بهترین روش­های رتبه­بندی به شمار می­رود که طبق آزمایشات انجام شده دارای دقت بالایی می­باشد. فرمول احتمالی BM25 که نشان دهنده شباهت میان پرس­وجوی Q با سند D می­باشد در رابطه 2-2 نشان داده شده است (جمع وزن همه واژه­ها در پرس­وجو).

|  |  |
| --- | --- |
| رابطه 2-2 |  |

k1 و b به صورت پیش­فرض 2/1 و 75/0 مقداردهی می­شوند. اما مقادیر کوچکتر b بعضی وقتها سودمندتر می­باشد. برای پرس­وجوهای با طول بزرگ، k3 اغلب مساوی 7 یا 1000 مقدار دهی میشود و k2 اغلب صفر است.

### 2-3-2- بازنمایی­های بردار متراکم

جاسازی کلمات[[46]](#footnote-46)، بردارهای عددی هستند که نمایانگر کلمات یک لغت نامه­اند و کاربردهای گسترده­ای در حوزه پردازش زبان طبیعی دارند (البته محدود به این حوزه نیستند!).

مفهوم اصلی جاسازی کلمات این است که تمامی لغات استفاده شده در یک زبان را می­توان توسط مجموعه­ای از اعداد اعشاری (در قالب یک بردار) بیان کرد. جاساز کلمات، بردارهای n بعدی­ای هستند که تلاش می­کنند معنای لغات و محتوای آنها را با مقادیر عددی خود ثبت و ضبط کنند. هر مجموعه­ای از اعداد یک "بردار کلمه" معتبر به حساب می­آید که الزاماً سودمند نیست، آن مجموعه­ای از بردار کلمات سودمندند که معنای کلمات، ارتباط بین آنها و محتوای کلمات مختلف را همان طور که به صورت طبیعی مورد استفاده قرار گرفته­اند، بدست آورده باشند. در ادامه به معرفی دو جاساز کلمات one-hot و word2vec پرداخته می­شود.

**الف) جاساز کلمات one-hot**

جاساز کلمات one-hot [194] طرح ساده­ای دارد و در آن هر المان در بردار به یک کلمه در لغت­نامه اختصاص داده می­شود. یعنی هر درایه این بردار معرّف یک کلمه از لغت­نامه خواهد بود. برای هر کلمه نیز تنها درایه متناظر برابر با ۱ بوده و مابقی درایه­ها با صفر مقداردهی خواهند شد. مثلا فرض کنید لغت­نامه ما فقط حاوی ۵ کلمه “پادشاه”، “ملکه”، “مرد”، “زن” و “بچه” باشد. اگر از شیوه­ای که تازه مطرح کردیم استفاده کنیم کلمه “ملکه” به صورت زیر بازنمایی خواهد شد:

پادشاه

مرد

0

1

0

0

0

بچه

زن

ملکه

شکل 2-2- نمونه­ای از جاسازی one-hot

با استفاده از این شیوه، نمی­توان هیچ قیاس معناداری بین دو بردار داشت و تنها می­توان مساوی بودن بردارها را بررسی کرد. نمونه­ای از جاسازی one-hot در شکل 2-3 قابل مشاهده است.

**ب) جاسازی کلمات word2vec**

در الگوریتم word2vec [195]، از بازنمایی توزیع شده برای هر کلمه استفاده می­شود. به عنوان مثال برداری با چند بُعد را در نظر بگیرید. هر کلمه توسط توزیعی از مقادیر عددی (وزن­ها) بر روی درایه­های مختلف بردار بازنمایی می­شود. بنابراین بجای نگاشت یک به یک بین یک درایه در بردار و یک کلمه در لغت نامه، بازنمایی یک کلمه در تمامی درایه­های یک بردار پخش می­شود، و هر درایه بردار در مشخص شدن معنای تعداد زیادی از کلمات نقش ایفا می­کند. برای درک بهتر این قضیه مثال شکل 2-3 مفید می­باشد. در این مثال کلمات مختلف توسط مقادیر مختلف هر درایه توصیف شده­اند و تمامی درایه­ها در ثبت و ضبط معنای کلمات مختلف نفش ایفا می­کنند. (در اینجا خصایص ویژه­ای به هر درایه اختصاص داده شده است. مثلاً درایه اول مشخص کننده صفت "حکمرانی"، درایه دوم بردار مشخص کننده صفت "مردانگی" و الی اخر است. در بردار کلمه پادشاه می­بینیم که درایه­های مختلف به ترتیب بر اساس میزان سنخیت با صفت با کلمه مورد بحث، دارای مقادیر(وزن­های) مختلفی­اند. بعنوان مثال صفت مردانگی برای پادشاه بسیار بالا بوده اما صفت زنانگی مقدار پایینی دارد این مساله برای ملکه و زن برعکس است.)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | **پادشاه** |  | **ملکه** |  | **زن** |  | **شاهزاده** |
| **حکمرانی** |  |  |  | **99/0** |  | **99/0** |  | **02/0** |  | **98/0** |
| **مردانگی** |  |  |  | **99/0** |  | **05/0** |  | **01/0** |  | **02/0** |
| **زنانگی** |  |  |  | **05/0** |  | **93/0** |  | **999/0** |  | **94/0** |
| **سن** |  |  |  | **7/0** |  | **6/0** |  | **5/0** |  | **1/0** |
| **...** |  | **...** |  | **...** |  | **...** |  | **...** |  | **...** |

شکل 2-3- نمونه­ای از جاسازی کلمات word2vec

چنین برداری بوسیله روشی انتزاعی، معنای یک کلمه را ارائه می­کند. صرفاً با استفاده از یک مجموعه متنی بزرگ، می­توان بردار کلماتی را فرا گرفت که قادر به ثبت و ضبط روابط بین کلمات باشند. نمونه­ای از جاسازی کلمات word2vec درشکل 2-4 قابل مشاهده است.

**ج) روش GloVe** [[47]](#footnote-47)

روش Glove [196]، یک روش جاسازی کلمات می­باشد که در آن می­توان روابط معنایی بین کلمات را از ماتریس هم­وقوعی[[48]](#footnote-48) به دست آورد. روال کار به صورت زیر است:

فرض کنید پیکره­ای با V کلمه در دسترس است. در گام اول، ماتریس هم وقوعی X با ابعاد V\*V ساخته می­شود. به نحوی که هر درایه (j و i) این ماتریس، نشان دهنده تعداد هم­وقوعی کلمات i و j می باشد. برای نمونه ماتریس هم­وقوعی پیکره The cat sat on the mat به صورت شکل 2-5 می­باشد.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mat | On | Sat | cat | The |  |
| 1 | 1 | 0 | 1 | 0 | The |
| 0 | 0 | 1 | 0 | 1 | cat |
| 0 | 1 | 0 | 1 | 0 | Sat |
| 0 | 0 | 1 | 0 | 1 | On |
| 0 | 0 | 0 | 0 | 1 | Mat |

شکل 2-4- ماتریس هم­وقوعی برای جمله  “the cat sat on the mat”

Xi,j : تعداد هم­وقوعی کلمات i و j.

Xi, j = Xj, i :تنها در صورتی این رابطه برقرار است که پنجره ای که برای context در نظر گرفته می­شود، n لغت قبل و بعد را شامل شود. مثلا یک لغت قبل وبعد، یا 4 لغت قبل و بعد.

فرمول مدل Glove در رابطه 2-3 قابل مشاهده است.

|  |  |
| --- | --- |
| رابطه 2-3 |  |

در این رابطه:

ej :نشان­دهنده بردار embadding لغتی که اندیس jدارد.

θi : وزن پارامترهای نورون iهست.

f(Xi‚j) : وزن که در ادامه بیشتر به آن پرداخته می ­شود.

* هدف به دست اوردن وابستگی بین iو j است که log آن برابر تعداددفعاتی است که i و j با هم دیده شده­اند.
* در ابتدا هم i و j به ازای تمام کلمات minimize می­شوند.
* bi و bj هم بایاس هستند.

کاربرد f(Xi‚j) :

ممکن است دوکلمه ی i و j با هم دیده نشده باشند پس Xi‚j=0 می­شود و در نهایت log(0) تعریف نشده است. بنابراین، دیگر عبارت محاسبه نمی­شود و مقدار صفر خواهد گرفت. پس برای حل مشکل تعریف نشده بودن f(Xi‚j) = 0 می­شود.

با وجود f(Xi‚j) برای کلمات پرتکرار وزن بیشتری نسبت داده می­شود و لذا، احتمال انتخاب شدن آنها افزایش یافته و همچنین احتمال انتخاب شدن کلمات کم­تکرار هم کنترل می­شود.

### 2-3-3- بازنمایی به روش برت[[49]](#footnote-49)

زمانی که برای اولین بار یک شبکه کانولوشنی در مسابقه Image Net برنده شد، توجهات به مبحث یادگیری ماشین و در ادامه یادگیری عمیق جلب شد و همگان راه حل تمام مشکلات را در این زمینه جستجو می‌کردند؛ ولی فراموش می‌کردند که مسابقهImage Net  داده عظیمی در اختیار شرکت کنندگان قرار می‌دهد و برگ برنده شبکه­های عمیق نیز همین داده زیاد است. در صورتی که برای خیلی از مشکلات این حجم از اطلاعات در دسترس نیست. از طرفی آموزش یک شبکه عمیق با داده زیاد، کار بسیار دشواری است. زیرا این کار نیاز به قدرت پردازشی زیادی دارد.

در این وضعیت به منظور مرتفع ساختن مشکل، استفاده از مدل­های پیش آموزشی به کمک افرادی آمد که از داده و قدرت پردازشی محدودی برخوردار بودند. شبکه­هایی که با داده مسابقه Image Netآموزش داده شده­اند، در اختیار همگان قرار دارد و دیگر نیازی نیست که تمام مسیر از ابتدا طی شود. کافیست که یک شبکه انتخاب شود و با استفاده از دو عمل استخراج ویژگی[[50]](#footnote-50) و بهبود پارامتر[[51]](#footnote-51) ، شبکه برای کار مورد نظر اختصاصی شود.

در سال ۲۰۱۸ این مسیر برای مسأله­های متنی یا به طور دقیق‌تر NLP  نیز در دسترس قرار گرفت. شبکه بزرگی با داده­های زیاد (Wikipedia + [BookCorpus](http://yknzhu.wixsite.com/mbweb))  توسط مهندسان گوگل آموزش داده شد و در دسترس همه قرار گرفت.  این شبکه برت نام دارد.

شبکه برت [40] در دو اندازه متفاوت آموزش داده شده­اند. برت پایه شامل ۱۲ لایه انکودر[[52]](#footnote-52) (که در Transformer Blocks نامیده می‌شوند) و شبکه بزرگ‌تر شامل ۲۴ لایه انکودر می­باشد.

هر انکودر از یک لایه خودتوجهی، یک لایه dense و یک لایه dropout تشکیل شده است. قبل از انکودرها لایه Embedding و بعد از آنها لایه­های خروجی قرار دارند. شبکه پایه در مجموع ۱۱۰ میلیون پارامتر و شبکه بزرگ ۳۴۵ میلیون پارامتر دارد.

برای آموزش برت از دو روش زیر استفاده می­شود:

الف) Masked Language Model (MLM)

در ابتدا ۱۵٪ لغات متن با نشانه [MASK] جایگزین شده و به ورودی برت داده می‌شوند. در این روش باید لغات جایگزین شده با نشانه MASK، براساس زمینه ارائه شده در متن، توسط شبکه حدس زده شود. آموزش این شبکه دو سویه (Bidirectional) است و به کلمات قبلی و بعدی حساس می­باشد. شمای کلی این روش در شکل 2-6 قابل رویت است. از نظر فنی، پیش­بینی کلمات خروجی مستلزم موارد زیر است:

الف) افزودن یک لایه طبقه­بندی در بالای خروجی رمزگذار.

ب) ضرب بردارهای خروجی با ماتریس جاسازی شده، تبدیل آنها به بعد واژگان.

ج) محاسبه احتمال هر کلمه در واژگان با softmax.

تابع هزینه در برت، تنها پیش­بینی­های MASK را در نظر می­گیرد و پیش­بینی کلمات بدون MASK را نادیده می­گیرد. در نتیجه، مدل کندتر از مدل­های جهت­دار همگرا می­شود.

|  |  |
| --- | --- |
|  |  |
| برت بزرگ | برت پایه |

شکل2-5- اندازه­های متفاوت برت پایه و برت بزرگ

ب) Next Sentence Prediction (NSP)

در فرآیند آموزش برت، مدل، جفت­های جملات را به عنوان ورودی دریافت می­کند و یاد می­گیرد پیش­بینی کند که آیا جمله دوم در این جفت، جمله بعدی در سند اصلی است. در طول آموزش، 50 درصد از ورودی­ها جفت­هایی هستند که در آنها جملات دوم، عبارت بعدی در سند اصلی است، در حالی که در 50 درصد دیگر یک جمله تصادفی از پیکره به عنوان جمله دوم انتخاب می­شود.

بازنمایی برت روشی است که بازنمایی را به صورت پویا و بر اساس کلمات اطراف در متن تولید می­کند. در حالی که بازنمایی­هایی مانند Word2Vec، هر کلمه را بدون در نظر گرفتن زمینه ای که در آن ظاهر شده است، تولید کرده و یک نمایش ثابت دارد. به عنوان مثال، دو جمله آورده شده است:

"شیر می­تواند بسیاری از حیوانات را بخورد."

"بسیاری از حیوانات می­توانند شیر بخورند."

Embedding to vocab + softmax

W’1

W’2

W’3

W’4

W’5

لایه طبقه­بندی: لایه کاملاً متصل + GELU + نرمالیزه

W5

W4

W3

W2

W1

Transformer Encoder

Embedding

W4

W3

W1

W2

W5

W’4

W’3

W4

W3

W5

W1

W2

شکل 2-6- آموزش برت از طریق MLM

در هر دو جمله، Word2Vec یک کلمه بازنمایی یکسان را برای کلمه "شیر" ایجاد می­کند ، در حالی که در برت، بازنمایی کلمه "شیر" برای هر جمله متفاوت است. جاسازی کلمات با توجه به زمینه، اَشکال دیگری از اطلاعات را در بر می­گیرد که به نمایش دقیق­تر ویژگی­ها منجر می­شود ، که به نوبه خود منجر به عملکرد بهتر مدل می­شود.

ورودی­های برت می­بایست در یک قالب مشخص، با نشانه­ ویژه برای شروع [CLS] و نشانه ­تفکیک/پایان جملات [SEP] مشخص شوند. سپس جملات به نشانه­هایی که با واژگان برت[[53]](#footnote-53) تطابق دارد، تبدیل می­شوند. در نهایت به هر یک از نشانه­های جملات یک id منحصر به فرد تعلق می­گیرد. شمای کلی آموزش برت به روش NSP در شکل 2-7 قابل مشاهده است.

2) اضافه کردن توکنهای [CLS] و [SEP]

**شیر**

می­تواند

**بسیاری**

**از**

**حیوانات**

**را**

**بخورد**

[SEP]

[CLS]

102

**2293**

**2006**

**12758**

**19379**

**14726**

**17453**

**1037**

101

3) جایگزینی توکنها با id مربوطه

**بخورد**

**را**

**حیوانات**

**از**

**بسیاری**

می­تواند

**شیر**

1) شکستن کلمات به توکنها

توکن گذاری

"شیر می­تواند بسیاری از حیوانات را بخورد"

شکل 2-7- آموزش برت به روش NSP

## 2-4- پارامترهای ارزیابی

بحث ارزیابی در سیستم­های پرسش و پاسخ، به عنوان یک چالش مهم در این نوع سیستم­ها، بسیار حائز اهمیت است. در غالب سیستم­های پرسش و پاسخ معیار دقت، به عنوان معیاری قابل قبول، مورد تایید می­باشد. در سیستم­های پرسش و پاسخ مدرن که اغلب مبتنی بر درک مطلب ماشین هستند، معیارهای ارزیابی دیگری از قبیل F1 و EM[[54]](#footnote-54) برای ارزیابی کارایی مدل مطرح می­باشد. همچنین از آنجایی که پاسخگویی به پرسش­ها محدود به دامنه متون نیست، لذا از معیارهای ROUGE-L[[55]](#footnote-55) [51] و BLEU[[56]](#footnote-56) [52] نیز به صورت گسترده­ای استفاده می­شود. در ادامه شرح مفصلی از معیارهای ارزیابی نام برده شده ارائه می­شود.

### 2-4-1- معیار دقت

معیار دقت در پاسخ به صحت و سقم پاسخ­ها معمولاً برای ارزیابی پرسش­های کلوز[[57]](#footnote-57) و چند گزینه­ای استفاده می­شود. حالتی را درنظر بگیری که مجموعه پرسش­های Q = {Q1, Q2, …, Qm} که شامل m پرسش است را در اختیار داریم. چنانچه مدلی به بتواند به n پرسش پاسخ صحیح دهد، آنگاه دقت مدل به صورت رابطه 2-4 می­باشد.

|  |  |
| --- | --- |
| رابطه 2-4 |  |

### 2-4-2- معیار EM

معیار EM حالت خاصی از معیار دقت است که در آن، چناچه پاسخ سیستم برابر با پاسخ حقیقی بود مقدار 1 و در غیر این صورت 0 برگردانده می­شود.

### 3-4-2- معیار F1

این معیار بر اساس ماتریس اغتشاش[[58]](#footnote-58) عمل می­نماید. نمونه­ای از ماتریس اغتشاش در شکل 2-8 قابل رویت می­باشد.

معیار F1 از رابطه 2-5 بدست می­آید.

|  |  |
| --- | --- |
| رابطه 2-5 |  |

که در این رابطه مقادیر P و R به ترتیب از روابط 2-6 و 2-7 محاسبه می­شوند.

|  |  |
| --- | --- |
| رابطه 2-6 |  |
| رابطه 2-7 |  |

|  |  |  |  |
| --- | --- | --- | --- |
| پیش­بینی توسط الگوریتم | | |  |
| خیر | بلی |  | برچسب واقعی |
| FN | TP | بلی |
| TN | FP | خیر |

شکل 2-8- نمونه­ای از یک ماتریس اغتشاش

### 4-4-2- معیار ROUGE-L

یک معیار ارزیابی است که در ابتدا برای خلاصه سازی خودکار متون ایجاد شد. در این معیار، ارزیابی از طریق میزان همپوشانی بین خروجی مدل و پاسخ حقیقی انجام می­شود. ارزیابی­های Rouge مختلفی برای نیازهای مختلف ارزیابی وجود دارند که عبارتند از: ROUGE-N، ROUGE-L، ROUGE-W و ROUGE-S. "L" در ROUGE-L نشان دهنده طولانی ترین دنباله مشترک (LCS) است و ROUGE-L را می توان به صورت رابطه 2-8 محاسبه کرد:

|  |  |
| --- | --- |
| رابطه 2-8 |  |

### 5-4-2- معیار BLUE

BLEU به طور گسترده­ای برای ارزیابی عملکرد ترجمه مورد استفاده قرار می­گیرد. هنگامی که از این معیار در سیستم پرسش و پاسخ استفاده می­شود، میزان شباهت بین پاسخ­های پیش­بینی شده و پاسخ­های حقیقی را محاسبه می­کند. نقطه قوت این معیار اندازه گیری دقیق آن است، که به صورت رابطه 2-9 محاسبه می­شود.

|  |  |
| --- | --- |
| رابطه 2-9 |  |

در این رابطه hk(ci)، نشان­دهنده n-gram، Kام در پاسخ کاندید Ci است. به طور مشابه hk(ri)، تعداد وقوع n-gram مذکور در پاسخ حقیقی را نشان می­دهد. مقدار Pn(C,R) در زمان کوتاه بودن پاسخ، بیشتر است. لذا، چنین دقتی به تنهایی قادر به اندازه­گیری مناسب شباهت نیست. درنتیجه از یک ضریب جریمه برای کاهش آن استفاده می­شود که به صورت رابطه 2-10 محاسبه می­گردد.

|  |  |
| --- | --- |
| رابطه 2-10 |  |

در ادامه معیار BLEU به صورت رابطه 2-11 محاسبه می­گردد.

|  |  |
| --- | --- |
| رابطه 2-11 |  |

در این رابطه N، حداکثر طول n گرام­ها را نشان می­دهد. wn، برابر با مقدار .

ارزیابی BLEU نه تنها می­تواند شباهت بین پاسخ­های نامزدها و پاسخ­های واقعی را ارزیابی کند، بلکه خوانایی نامزدها را نیز آزمایش می­کند.

## **5-2- توابع هزینه[[59]](#footnote-59) مختلف مورد استفاده**

هنگام کار بر روی یک مساله یادگیری ماشین یا یادگیری عمیق، از توابع هزینه برای بهینه­سازی مدل در طول آموزش استفاده می­شود. می­توان گفت، هدف، تقریباً همیشه به حداقل رساندن تابع هزینه است. هرچه هزینه کمتر باشد، مدل بهتر است.

**الف) تابع هزینه cross-entropy** [197]

کلود شانون[[60]](#footnote-60) مفهوم انتروپی اطلاعات را در مقاله 1948 خود با عنوان "نظریه ریاضی ارتباطات" معرفی کرد [160]. به گفته شانون، آنتروپی یک متغیر تصادفی سطح متوسط "اطلاعات" ، "شگفتی" یا "عدم قطعیت" ذاتی در نتایج احتمالی متغیر است.

Cross-entropy را می­توان در دسته­بندی­های دوتایی یا چندتایی اعمال نمود.

* Binary Cross-entropy

از این تابع هزینه در حالتی که مقادیر هدف صفر یا یک هستند، استفاده می­شود. فرمول آن در رابطه 2-12 قابل مشاهده است.

|  |  |
| --- | --- |
| رابطه 2-12 |  |

* Categorical Cross-entropy

یکی دیگر از توابع هزینه بسیار مهم و کاربردی در مسائل طبقه­بندی، Categorical Cross-entropy است. این تابع هزینه، تابع پیش فرض برای مسائل طبقه­بندی چند کلاسی است که در آن به هر کلاس یک عدد صحیح منحصر به فرد بین 0 تا (1- تعداد کلاس­ها) نسبت داده می­شود. در این تابع، میانگین اختلاف توزیع احتمال واقعی و پیش­بینی شده برای همه کلاس­ها محاسبه می­شود. هرچه انتروپی پایین­تر باشد عملکرد مدل بهتر است. در بهترین حالت انتروپی صفر می­باشد. این تابع هزینه در رابطه 2-13 قابل مشاهده است.

|  |  |
| --- | --- |
| رابطه 2-13 |  |

در [16] از شبکه برت به عنوان انکودر اعمال شده به پرسش و متون استفاده شده است و برای آموزش آن از Loss به صورت رابطه 2-14 استفاده می­شود.

|  |  |
| --- | --- |
| رابطه 2-14 |  |

در این رابطه نشان دهنده پرسش، نشان دهنده متن مرتبط، و نشان دهنده متون غیر مرتبط می­باشد.

**ب) تابع هزینه log likelihood** [198]

یک روش آماری برای تخمین پارامتر‌(های) یک مدل، تابع تخمین گر بیشینه شباهت[[61]](#footnote-61) (MLE) می­باشد. ایده اصلی پشت MLE این است: مدلی که بهترین توصیف از داده را داشته باشد، بهترین پیش‌بینی‌کننده خواهد بود. مزیت اصلی MLE خاصیت تقریبی بودن آن است. و این یعنی زمانی که اندازه داده افزایش می‌یابد، تخمین با سرعت بیشتری به سمت پارامتر جمعیت همگرا می‌شود. از MLE برای تکنیک‌های مختلفی در آمار به منظور پیش‌بینی پارامترها استفاده می‌شود. در ادامه گام‌های کلی برای یافتن یک تخمین مناسب از پارامتر‌ها شرح داده شده است

گام ۱: ساخت یک فرض درباره تابع تولید داده.

گام ۲: فرموله کردن تابع درستی برای داده با استفاده از تابع تولید داده. تابع درستی چیزی به جز احتمال مشاهده داده با توجه به پارامتر(ها) مربوطه نیست. پارامترها به فرض‌ها و تابع تولید داده وابسته هستند.

گام 3:‌ پیدا کردن یک تخمین‌زن برای پارامتر(ها) با استفاده از تکنیک بهینه‌سازی. این کار با توجه به پیدا کردن تخمینی که تابع درستی(Likelihood Function) را بیشینه می‌کند انجام می‌شود.

متغير تصادفي با تابع چگالي تصادفي را درنظر بگیرید. محتمل­ترين مقدار x متناظر است با مقداري از x که موجب ماکزيمم شدن می­گردد. با توجه به این توضيح و با فرض اينکه مقادير مشاهده­هاي انجام شده را مي­توان به صورت z = g(s,v) بيان نمود، پيشنهاد مي­گردد که تخمين s با يافتن مقداري از s بدست آید که محتمل ترين مقدار آن براي توليد z است.

|  |  |
| --- | --- |
| رابطه 2-15 |  |

به عبارت دیگر

|  |  |
| --- | --- |
| رابطه 2-16 |  |

همچنين با توجه به اينکه تابع لگاريتم طبيعي يک تابع بطور يکنواخت افزايشي است، لذا مي­توان رابطه 2-16 را به صورت رابطه 2-17بازنويسي نمود و با عنوان log-likelihood function شناحته مي­شود.

|  |  |
| --- | --- |
| رابطه 2-17 |  |

**ج) تابع هزینه Negative log likelihood** [198]

یک تابع هزینه که کاربردهای فراوانی در حوزه دسته­بندی دارد. در واقع برخلاف توابع هزینه مذکور میزان بد بودن عملکرد را گزارش می­نماید. اکثر چارچوب های یادگیری ماشین، بهینه­سازی را از طریق حداقل­سازی خطا انجام می­دهند، اما در این تابع احتمال انتخاب دسته مناسب به حداکثر رسانده می­شود.

|  |  |
| --- | --- |
| رابطه 2-18 |  |

## **6-2- تقسیم­بندی سیستم­های پرسش و پاسخ**

می­توان سیستم­های پرسش و پاسخ را به طرق مختلف تقسیم­بندی کرد که سه تقسم­بندی مهم آن، عبارتند از:

* تقسیم­بندی بر اساس روش طراحی سیستم­های پرسش و پاسخ
* تقسیم­بندی بر اساس نوع منبع اطلاعاتی که پاسخ­ها از آن مشتق شده­اند.
* تقسیم­بندی سیستم­های پرسش و پاسخ بر اساس حوزه کارکرد.

که در ادامه به هریک از آنها پرداخته می­شود:

### 2-6-1- تقسیم­بندی بر اساس روش طراحی سیستم­های پرسش و پاسخ

سیستم­های پرسش و پاسخ از منظر روش طراحی، به دو دسته تقسیم­ می­شوند که عبارتند از:

* سیستم­های مبتنی بر شباهت ظاهری لغات
* سیستم­های مبتنی بر نقش کلمات در جمله

### 1-1-6-2- سیستم­های مبتنی بر شباهت ظاهری لغات

این سیستم­ها شامل روش­هایی است که عمدتاً بر اساس مجموعه کلمات[[62]](#footnote-62) عمل می­کنند. در واقع در نظر گرفتن پرسش به عنوان مجموعه­ای از کلمات، ساده­ترین روش پاسخگویی می­باشد. دبلیو بی کرُفت[[63]](#footnote-63) و همکاران در سال 2010 [53] یک سیستم پرسش و پاسخ مبتنی بر شباهت ظاهری لغات پیشنهاد کردند که در آن، ابتدا کلمات بی­معنی و ایست واژه­ها[[64]](#footnote-64) حذف شده و مابقی کلمات در مجموعه کلمات قرار می­گیرند. سپس جستجوی پاسخ بر اساس معیار شباهت (تعداد کلمات مشابه در پرسش و متون یا بخشی از متون) انجام می­شود. در این روش هیچ توجهی به ساختار و ارتباط بین کلمات وجود ندارند و از روش­­های خیلی ساده و ابتدایی برای تجزیه و تحلیل پرسش استفاده می­شود. با توجه به مطالب بیان شده، بدیهی است که اطلاعات مستخرج در این­گونه روش­ها نمی­تواند چندان دقیق باشد.

### 2-1-4-2- سیستم­های مبتنی بر نقش کلمات در جمله

این سیستم­ها از ابزارهای مورد استفاده در متن­کاوی و پردازش زبان طبیعی به طور گسترده استفاده می­کنند و خود به چندین دسته تقسیم می­شوند که در ذیل به آن پرداخته می­شود.

**الف) سیستم­هایی که بر مبنای آنالیز نحوی و ریخت­شناسی عمل می­کنند.**

چنین سیستم­هایی به طور مرسوم با به کارگیری ابزارهای مختلف برای آنالیز نحوی و ریخت شناسی[[65]](#footnote-65) متون سعی در یافتن روابط ساختاری بین کلمات در جملات دارند. کلومیتس[[66]](#footnote-66) و همکارش در سال 2011 [1] اَشکال مختلف هر کلمه را به عنوان یک واحد در نظر گرفتند. آنها با با بهره­گیری از ابزارهای stemmer و lemmatizer سعی در افزایش احتمال یافتن پاسخ دقیق­تر را داشتند. این کار از یک سو مانع ظهور حالت­های مختلف یک کلمه یا عبارت در متن می­شود و از سوی دیگر، باعث از بین رفتن برخی اطلاعات مانند زمان فعل می­گردد. آنها از ابزار برچسب گذار اجزای کلام[[67]](#footnote-67) به منظور تجزیه و تحلیل نحوی زبان طبیعی و از ابزارهای چانکر[[68]](#footnote-68) و پارسر[[69]](#footnote-69)، برای یافتن گروه­های اسمی و فعلی استفاده نمودند.

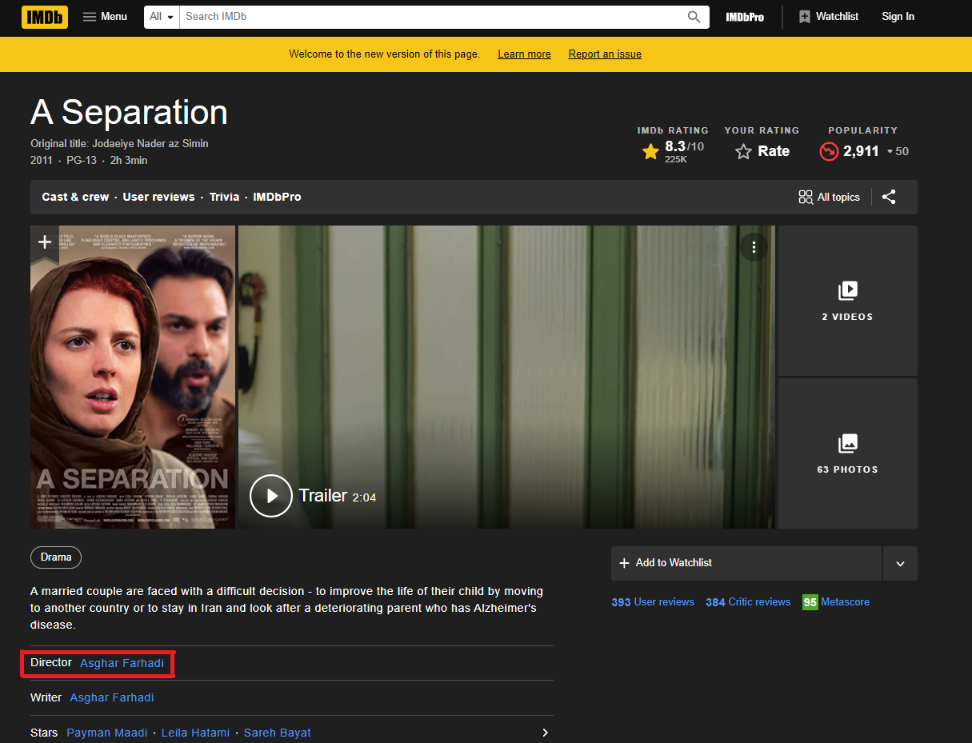
با استفاده از پارسر، روابط معنایی بین جملات و پرسش استخراج شده و پاسخ به عنوان خروجی سیستم برگردانده می­شود [55]. روابط ساختاری بین کلمات در جمله و پرسش، از طریق تجزیه و تحلیل ریختی جملات بهتر مشخص می­شود. در نظر گرفتن ویژگی­های نحوی منجر به افزایش کارایی نزدیک به 10 درصدی در سیستم­های پرسش و پاسخ، می­شود [56]. نکته قابل تأمل این است که اعمال تجزیه و تحلیل­های بیان شده، پیچیدگی محاسباتی را افزایش می­دهد.

**ب) استفاده از منطق مرتبه اول برای پردازش پرسش­ها و اسناد**

معمولاً به منظور تولید گزاره­های منطقی از عبارات زبان طبیعی، از شناسایی نقش­های معنایی لغات استفاده می­شود. مولدوان[[70]](#footnote-70) و همکاران در سال 2007 [57]، به منظور تولید گزاره­های مرتبه اول در پایگاه داده رابطه­ای از ابزاری به نام MRL[[71]](#footnote-71) استفاده کردند. در این روش، به عنوان مثال برای پاسخگویی به پرسش "چه کسی حسین را کشت؟" ابتدا، پرسش با استفاده از MRL به صورت [(حسین و x)کشت] تبدیل می­شود که در آن پارامتر اوّل، قاتل و پارامتر دوّم قربانی را بیان می­کند. x، متغیر است و برنامه منطقی به دنبال نمونه­ای می­گردد که فعل کشت با مقدار مشخص حسین برای پارامتر دوم بیاید. از آنجایی که این روش، دارای حجم زیادی از محاسبات پرهزینه برای ادراک عمیق پرسش­ها و پاسخ­های بالقوه می­باشد، لذا در سیستم­های پرسش و پاسخ کمتر به آن پرداخته شده است. اما این روش، به عنوان نقطع عطفی برای سیستم­های پرسش و پاسخ آینده بحساب می­آید.

**ج) سیستم­های مبتنی بر روابط معنایی**

در این روش، پاسخ­ها از روابط معنایی موجود در متن استخراج می­شود. دالی[[72]](#footnote-72) و همکاران در سال 2009 [55] یک سیستم پرسش و پاسخ مبتنی بر نظریه گراف را پیشنهاد می­کنند. در روش پیشنهادی آنها به ازای هر جمله یک رابطه فاعل، فعل و مفعول استخراج می­شود و سپس با استفاده از نظریه گراف، پاسخ­ها استخراج می­شوند. به عنوان نمونه دیگر، کاتز[[73]](#footnote-73) و همکاران در سال 2002 [58] سیستمی به نام استارت[[74]](#footnote-74) را معرفی کردند که در آن اطلاعات با استفاده از تعدادی الگوهای از قبل تعیین شده استخراج و در بانک­های اطلاعاتی مجزا دخیره می­شوند. برای نمونه، برای پرسش "Who directed a separation?" پاسخ نام یک کارگردان است و این پرسش می­بایست در بانک اطلاعاتی مربوط به فیلم­ها و کارگردان­ها اعمال گردد. پرسش به سه تایی (get “IMDB-movie” “A Separation?” “DIRECTOR”) تبدیل می­شود. نکته مهم در این روش این است که اطلاعات استخراج شده در این سیستم محدود به دانش طراح سیستم می­باشد. به عنوان مثال، در شکل 2-9 محل قرارگیری اطلاعات مربوط به کارگردان از قبل مشخص شده است. به عنوان مثال دیگر، چنانچه پرسش در مورد اسم شخص به خصوصی باشد، پاراگراف اول ویکی­پدیا به عنوان پاسخ برگردانده می­شود.



شکل 2-9- استخراج نام کارگردان فیلم A Separation با استفاده از سیستم استارت

­ **د) سیستم­های ترکیبی**

در این سیستم­ها از تکنیک­های خلاصه­سازی در کنار روش­های مرسوم پرسش و پاسخ استفاده می­شود. به بیان بهتر، در این سیستم­ها استفاده از تکنیک خلاصه­سازی به دو دلیل انجام می­شود. نخست، به منظور خلاصه­سازی متن اصلی، دوم، برای کوتاه کردن پاسخ­های استخراج شده [58]، [59]، [60] که در ذیل به آنها می­پردازیم.

* **استفاده از تکنیک خلاصه­سازی برای خلاصه کردن متن اصلی**

در این حالت، سیستم از خلاصه ایجاد شده، پاسخ­ها را استخراج می­کند. بیریوکُو[[75]](#footnote-75) و همکاران در سال 2005 [61]، میزان تاثیر عناوین اسناد در خلاصه­سازی چند سنده برای انواع خاصی از پرسش­ها را بررسی کردند. برای نمونه، چنانچه پرسش در مورد افراد است و پرسش به صورت Who is Z? پرسیده شود. آنگاه Z یک فرد است.

دیمر فاشمن[[76]](#footnote-76) و همکارش در سال 2006 [62]، یک سیستم پرسش و پاسخ در حوزه درمان و با استفاده از روش ترکیبی ارائه کردند. به عنوان نمونه­ای از عملکرد این سیستم، برای پرسش­ What is the best treatment for cancer?، ابتدا روش­های درمانی را از یک مجموعه معتبر و استاندارد­ مشخص می­کند و آنها را بر اساس ارتباط­شان دسته­بندی می­کند. در مرحله بعد خلاصه­ای از آنها تهیه می­گردد که برای پاسخگویی به پرسش­ها، اطلاعات مناسبی در اختیار می­گذارد. خلاصه­ی ایجاد شده شامل عنوان، چکیده و جملات با بیشترین امتیاز می­باشد.

شی[[77]](#footnote-77) و همکاران در سال 2007 [63] روش BioSquash را برای پاسخگویی به پرسش­ها پیشنهاد کردند که با بکارگیری خلاصه­سازی چند سنده به پرسش­ها پاسخ می­دهد. خلاصه ایجاد شده در این روش دارای 4 جزء اصلی است که عبارتند از: Annotatior، که پرسش­ها و مستندات را به صورت نحوی تفسیر می­کند. تشابه مفهومی[[78]](#footnote-78)، که اطلاعات معنایی سطحی و ارتباطات بین مفاهیم در پرسش­ها و مستندات را مشخص می­کند. ماژول استخراج­گر، که انتخاب محتوی را انجام می­دهد و ماژول ویرایش­گر، که بر خوانایی ادبی متمرکز است.

* **سیستم­هایی که برای پاسخگویی از خلاصه متن استفاده می­کنند**

تورس مورنو و همکاران در سال 2009 [60]، سیستم جدیدی از ترکیب دو سیستم خلاصه­سازی عمومی و پرسش و پاسخ پیشنهاد کردند. خلاصه­سازی عمومی بخش­های مهم متن را مشخص می­کند. ادعای این روش این است که استفاده از سیستم خلاصه­سازی عمومی چند سنده به همراه چندین پارامتر فشرده­سازی در یک سیستم پرسش و پاسخ باعث کاهش فضای جستجو و در نتیجه افزایش پاسخ­های صحیح می­شود.

### 2-3-2- تقسیم­بندی بر اساس نوع منبع اطلاعاتی که پاسخ­ها از آن مشتق شده­اند.

از منظر نوع منبع اطلاعاتی، سیستم­های پرسش و پاسخ به دو دسته مبتنی بر اسناد و مبتنی بر پایگاه دانش تقسیم­بندی می­شوند. سیستم­های پرسش و پاسخ مبتنی بر اسناد، از اسناد متنی بدون ساختار برای پاسخگویی استفاده می­کند، در حالی که سیستم­های پرسش و پاسخ مبتنی بر پایگاه دانش[[79]](#footnote-79) از پایگاه­های دانش ساختار یافته و از پیش تعریف شده که اغلب به صورت دستی ساخته می­شوند، برای پاسخگویی استفاده می­کنند. سیستم­های پرسش و پاسخ متنی در مقایسه با سیستم­های پرسش و پاسخ مبتنی بر پایگاه دانش بسیار مقیاس­پذیرتر هستند، زیرا منابع متنی بدون ساختار که برای به دست آوردن پاسخ، مورد استفاده قرار می­گیرند نسبتاً رایج بوده و به راحتی در دسترس هستند. مانند: ویکی­پدیا، اخبار، مقالات، کتب علمی و غیره. در حالی که پایگاه­های دانش، نیاز به صرف زمان و منابع مختلف زیادی برای تولید دارند.

### 2-3-3- دسته­بندی سیستم­های پرسش و پاسخ بر اساس حوزه کارکرد

به دلیل توسعه روز افزون سیستم­های پرسش و پاسخ و گستردگی زیاد آن دسته­بندی­های مختلفی را می­توان بر روی آن انجام داد. یکی از دسته­بندی­های متداول و مؤثر، دسته­بندی بر اساس حوزه کارکرد آنها می­باشد. برخی از سیستم­ها تنها در یک حوزه معین و مشخص توانایی پاسخگویی دارند. برای مثال سیستم­هایی که توانایی پاسخگویی در مسائل پزشکی را دارند و در دیگر حوزه­ها ناتوانند. اما دسته دیگری از سیستم­ها وجود دارند که محدود به حوزه خاصی نیستند و توانایی پاسخگویی به پرسش­ها را به صورت عمومی دارند. بناراین می­توان سیستم­های پرسش و پاسخ را بر اساس حوزه کارکردشان به دو دسته تقسیم نمود.

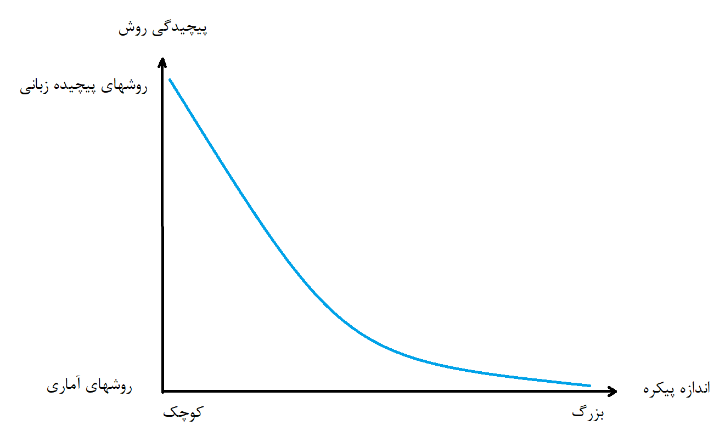
* سیستم­های پرسش و پاسخ با دامنه محدود[[80]](#footnote-80)
* سیستم­های پرسش و پاسخ با دامنه باز[[81]](#footnote-81)

سیستم­های پرسش و پاسخ دامنه باز و دامنه محدود به دلایل مختلفی متمایز هستند که در ذیل به آن می­پردازیم.

**الف) حجم داده­ها:**

یکی از مهمترین جنبه­های تمایز بین سیستم­های پرسش و پاسخ دامنه باز و دامنه بسته، میزان داده­های در دسترس برای پاسخگویی به پرسش­ها است. از آنجایی که در سیستم­های پرسش و پاسخ دامنه باز داده­های بیشتری در دسترس است، لذا از روش­های متفاوتی برای پاسخگویی به پرسش­ها استفاده می­شود. بریل[[82]](#footnote-82) و همکاران در سال 2001 [54]، یک روش مبتنی بر افزونگی[[83]](#footnote-83) برای سیستم­های پرسش و پاسخ حوزه باز پیشنهاد کردند. بریل دریافت هرچه اندازه پیکره[[84]](#footnote-84) بزرگ­تر باشد، احتمال یافتن پاسخ یک پرسش از طرق روش­های داده­محور[[85]](#footnote-85) بیشتر است و دیگر نیازی به استفاده از مدل­های پیچیده زبانی نیست. استفاده از روش مذکور در سیستم­هایی مناسب است که، حجم داده­های موجود در مجموعه اسناد آنها قابل توجه بوده و افزونگی مناسبی در این داده­ها وجود داشته باشد. این حالت، غالباً در سیستم­های پرسش و پاسخ با دامنه باز اتفاق می­افتد. از سوی دیگر تکنیک­های دیگری برای استخراج پاسخ وجود دارند که مبتنی بر عملیات­های پیچیده زبان طبیعی مانند استنتاج[[86]](#footnote-86) می­باشند. علی­رغم دقت قابل قبول این تکنیک­ها، به دلیل پیچیدگی محاسباتی بالا اکثراً در سیستم­های دامنه محدود به کار گرفته می­شوند که حجم داده و افزونگی کمتری در آنها وجود دارد [64].

با توجه به مطالب بیان شده، می­توان نتیجه گرفت که در سیستم­های پرسش و پاسخ بین حجم داده­های موجود و پیچیدگی تکنیک­های مورد استفاده، می­بایست توازن وجود داشته باشد. هرچه حجم داده­ها بیشتر باشد استفاده از تکنیک­های داده­محور که پیچیدگی کمتری دارند و بالعکس هرچه حجم داده­ها کمتر باشد بکارگیری تکنیک­های پیچیده­تری مانند پردازش زبان طبیعی مقرون به صرفه است. رابطه بین حجم داده­ها و پیچیدگی روش­های پردازش در شکل 2-10 نشان داده شده است.



شکل 2-10- رابطه بین اندازه داده­ها و پیچیدگی روش­های پردازش

**ب) بافت دامنه**[[87]](#footnote-87)**:**

ناویگلی[[88]](#footnote-88) در سال 2009 [65] در مقاله خود تحت عنوان ابهام­زدایی لغات، بافت دامنه را یک عامل بسیار مهم تمایز در سیستم­های پرسش و پاسخ دامنه باز و دامنه بسته می­داند. وی در این مقاله توضیح می­دهد که هر چه دامنه گسترده­تر شود مشکلات جدیدتری بروز می­کند. برای مثال بعضی کلمات وجود دارند که شکل نوشتاری یکسانی دارند ولی مفهوم آنها با یکدیگر متفاوت است (برای مثال، کلمه "شیر" در عبارت شیر جنگل و شیر خوراکی). این ابهام به وسیله بافتی که کلمه در آن قرار دارد، مرتفع می­گردد. برای این منظور از تکنیک­های رفع ابهام از کلمه[[89]](#footnote-89) استفاده شده است. این تکنیک­ها زمانی کارآمد هستند که دامنه خیلی گسترده نباشد. هرچه دامنه گسترده­تر باشد، احتمال پدیدار شدن کلمات با معنی متفاوت بیشتر می­شود و در نتیجه فرآیند رفع ابهام دشوارتر خواهد بود. از این رو سیستم­های دامنه باز نسبت به سیستم­های دامنه بسته نیازمندی بیشتری به تکنیک­های رفع ابهام از مفهوم کلمات دارند.

**ج) کاربران**

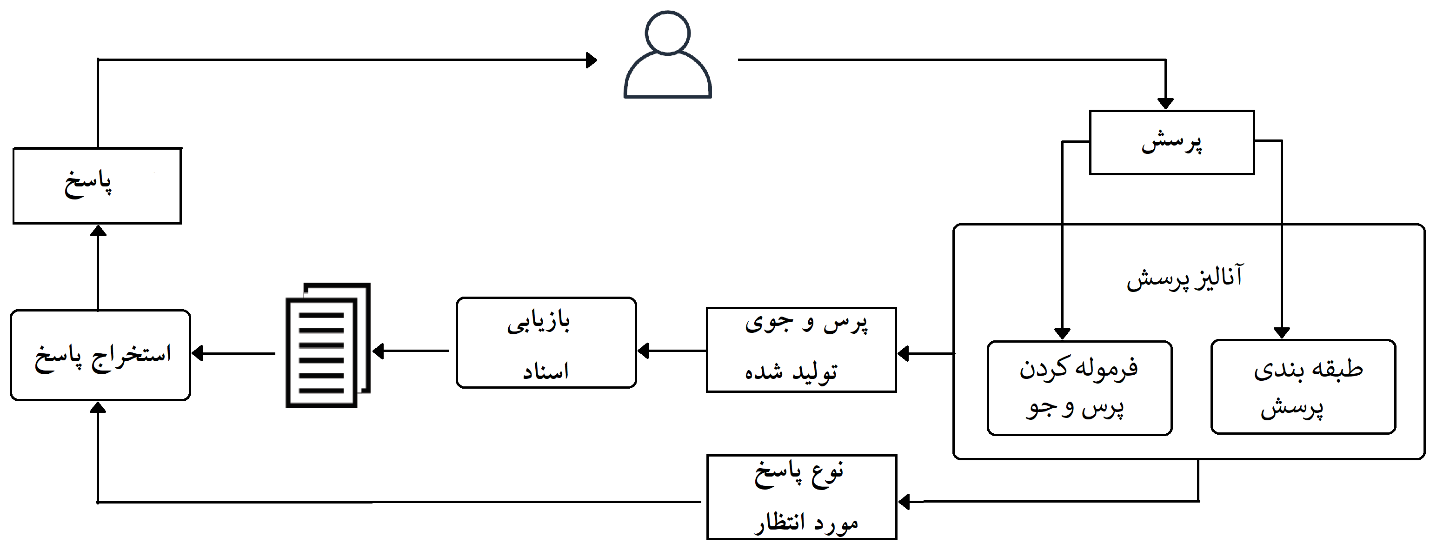
سیستم­های پرسش و پاسخ با دامنه باز و بسته با کاربران متفاوتی روبرو هستند. همین امر سبب می­شود که نوع پرسش­ها متفاوت باشند. کاربران در یک سیستم دامنه بسته، افراد خبره در آن حوزه هستند. پرسش­ها فنی­تر و تخصصی­تر می­باشند و معمولا از اصطلاحات تخصصی در پرسش­ها استفاده می­شود. پرسش­هایی که در این نوع سیستم­ها پرسیده می­شوند عموماً پیچیده­تر از پرسش­هایی است که در یک سیستم دامنه باز پرسیده می­شود.

**د) منابع اضافی[[90]](#footnote-90)**

تفاوت مهم دیگر سیستم­های پرسش و پاسخ دامنه باز و بسته امکان بهره­گیری سیستم­های دامنه بسته از منابع خاصی که معمولا برای کاربران آن دامنه توسعه یافته است و یا امکان ایجاد توافق بر سر یک نمادسازی استاندارد می­باشد که می­توان در فرآیند پاسخگویی به پرسش­ها از آن بهره جست.

### 2-4- معماری سنتی سیستم های پرسش و پاسخ

 معماری سنتی سیستم­های پرسش و پاسخ در شکل 2-11 نشان داده شده است [193]. که عمدتاً شامل سه مرحله است: آنالیز پرسش، بازیابی سند و استخراج پاسخ [67] و [68]. با توجه به این نکته که پرسش به زبان طبیعی مطرح می­شود، هدف از تجزیه و تحلیل پرسش، درک مناسب و مؤثر پرسش است. چراکه درک درست از پرسش منجر به بازیابی آسان­تر اسناد حاوی پاسخ و استخراج مؤثرتر و دقیق­تر پاسخ­ها خواهد شد. بنابراین عملکرد مناسب این مرحله تاثیر قابل توجهی بر مراحل بعدی دارد و لذا برای خروجی نهایی سیستم مهم است [70]. در مرحله بازیابی اسناد، اسناد مرتبط با پرسش بر اساس یک سیستم بازیابی اسناد [69] یا یک موتور جستجوی وب [17] و [18] به دست می­آیند و در اختیار مرحله استخراج پاسخ قرار می­گیرند. در نهایت،  در مرحله استخراج پاسخ، پاسخ­های دقیق و مرتبط با پرسش ورودی از اسناد به دست آمده از مرحله قبل استخراج می­شود. در ادامه هر مرحله یک به یک تحلیل خواهند شد.



شکل 2-11- نمونه­ای از معماری سنتی در سیستم­های پرسش و پاسخ [193]

### 2-4-1- آنالیز پرسش

مرحله تجزیه و تحلیل پرسش دو هدف را دنبال می­کند.

الف) تسهیل بازیابی اسناد مرتبط با پرسش. لذا یک ماژول تدوین پرس­وجو به منظور تولید پرس­وجوهای مناسب در نظر گرفته می‌شود.

ب) طبقه­بندی و پیش­بینی نوع پرسش ورودی و به تبع آن نوع پاسخ مرتبط که سعی در بهبود عملکرد مرحله استخراج پاسخ دارد [73].

استرزالکوسکی[[91]](#footnote-91) و همکاران در سال 2006 [74]، روش تولید پرس­وجو از روی پرسش کاربر را به دو دسته تقسیم کرده­اند که شامل روش­های مبتنی بر دسته­بندی[[92]](#footnote-92) و روش­های غیر دسته­بندی[[93]](#footnote-93) است که در ذیل به هر یک پرداخته شده است.

**الف) روش­های مبتنی بر دسته­­بندی**

تاثیر دسته­بندی پرسش در افزایش دقت سیستم­های پرسش و پاسخ غیر قابل انکار است و نزدیک به 4/36 درصد خطاهای یک سیستم پرسش و پاسخ، به علت دسته­بندی نادرست پرسش است [75]. با وجود معماری­های متفاوت سیستم­های پرسش و پاسخ، بخش دسته­بندی پرسش جزء لاینفک همه آنهاست. لذا دسته­بندی پرسش نقش کلیدی در سیستم­های پرسش و پاسخ ایفا می­نماید [76]. در واقع به عملیات انتصاب یک بر چسب به پرسش داده شده به زبان طبیعی، دسته­بندی پرسش گفته می­شود. برای مثال در پرسش "کدام خیابان لندن محل زندگی روزنامه نگاران انگلیسی است؟" برچسب نسبت داده شده "مکان[[94]](#footnote-94)" می­باشد که نشان می­دهد پاسخ مورد انتظار از نوع یک محل جغرافیایی می­باشد. در واقع با استفاده از دسته­بندی پرسش می­توان نوع پاسخ مربوطه را مشخص کرد. به همین جهت به دسته­بندی پرسش پیش­بینی کننده نوع پاسخ گفته می­شود.

جدول 2-1- دسته­های پرسش ریز و درشت [77]

|  |  |
| --- | --- |
| **ریز** | **درشت** |
| abbreviation, expansion | ABBR |
| definition, description, manner, reason | DESC |
| animal, body, color, creation, currency, disease, event, food, instrument, language, letter, other, plant, product, religion, sport, substance, symbol, technique, term, vehicle, word | ENTY |
| Description, group, individual, title | HUM |
| city, country, mountain, other, state | LOC |
| Code, count, date, distance, money, order, other, percent, period, speed, temperature, size, weight | NUM |

تاکنون دسته­بندی­های متفاوتی برای پرسش­ها مطرح شده است. لی[[95]](#footnote-95) و روث[[96]](#footnote-96) در سال 2002 [77] دو دسته، با نام­های ریز[[97]](#footnote-97) و درشت[[98]](#footnote-98) برای دسته­بندی پرسش­ها پیشنهاد کرده­اند. این دسته­بندی­ها همپوشانی دارند و به صورت سلسله مراتبی عمل می­نمایند. دسته­بندی ریز 50 حالت و دسته­بندی درشت 6 حالت دارند. این دسته­بندی­ها در جدول 2-1 قابل مشاهده است.

به طور کلی می­توان سه هدف را برای دسته­بندی در نظر گرفت: کاهش فضای جستجو، مشخص کردن نوع پاسخ و انتخاب استراتژی جستجوی مناسب.

* **کاهش فضای جستجو**

چنانچه برچسب پرسش مشخص باشد، می­توان فضای جستجو را تقلیل داد و در بخش­هایی از متن که حاوی اطلاعات مربوط به برچسب هستند جستجو نمود. به عنوان نمونه، در مثال پیشین که بر چسب پرسش " کدام خیابان لندن محل زندگی روزنامه نگاران انگلیسی است؟ " از نوع مکان تعیین شد. می­بایست در متون و پارگراف­هایی که از مکان صحبت شده است، جستجو نمود.

* **مشخص کردن نوع پاسخ**

این ماژول وظیفه شناسایی نوع پرسش ورودی را به یکی از انواع (به عنوان مثال کجا، چه زمانی، چه کسی، چه چیزی ) و یا طبقه­بندی [78] و [79] که به صورت دستی توسط متخصصان حوزه زبان تعریف شده است را به عهده دارد. پس از اینکه نوع پرسش مشخص شد انواع پاسخ مورد انتظار را می­توان به راحتی و با استفاده از روش­های نگاشت مبتنی بر قانون تعیین کرد [11]. برای مثال پرسش "انقلاب اسلامی ایران چه زمانی اتفاق افتاد؟" چنانچه بدانیم نوع پرسش "چه زمانی" می­باشد، آنگاه نوع پاسخ "تاریخ" استنباط می‌شود. شناسایی نوع پرسش می­‌تواند محدودیت­هایی را در استخراج پاسخ ایجاد کند که این امر به طور قابل توجهی دشواری یافتن پاسخ­های صحیح را کاهش می‌دهد. طبقه­بندی پرسش در بسیاری مقالات [71] و [72] و [80] و [81] و [82] مورد توجه قرار گرفته است. برای مثال زائو[[99]](#footnote-99) و همکاران [82]، پیشنهاد کرده­اند که کلمات مرتبط از یک پرسش داده شده استخراج شوند و سپس پرسش بر اساس قوانین مرتبط با این کلمات استخراج شده به مفاهیم مختلف طبقه­بندی شود. همچنین کالینز[[100]](#footnote-100) و همکاران در سال 2003 [80]، با استفاده از تکنیک­های مختلف یادگیری ماشین مانند ماشین بردار پشتیبان[[101]](#footnote-101)، نزدیکترین همسایه­ها[[102]](#footnote-102) و درخت­های تصمیم­گیری[[103]](#footnote-103) لیستی از طبقه­بندی کننده­های پرسش را آموزش داده­اند.

* **انتخاب استراتژی جستجوی مناسب**

براساس نوع برچسب می­توان استراتژی مناسبی برای جستجو انتخاب کرد. به عنوان مثال در پرسش "صفحه نمایش 8K چیست؟" که برچسب آن از نوع تعریف[[104]](#footnote-104) می­باشد، می­توان استراتژی یافتن پاسخ را به صورت جملات از پیش تعریف شده­ای که با "صفحه نمایش 8K یک ..." شروع می­شود، در نظر گرفت.

به نظر می رسد ساده­ترین راه برای دسته­بندی پرسش­ها استفاده از قواعد (دستی) از پیش تعریف شده است. قوانین مورد استفاده می تواند ساده باشد مثل "اگر پرسشی که با چه کسی شروع شود به کلاس شخص دسته­بندی می شود" و یا "اگر پرسشی که با کجا شروع شود با پرچسب مکانی برچسب زده می‌شود" و یا ... . همچنین قواعد می­تواند پیچیده باشد و همراه با حالت برچسب­گذاری شده، تجزیه شده و یا معنایی استفاده شود [83].

اما این روش­ها مزایا و معایب خاص خود را دارا می­باشند.

* محاسبات سریع و ارزان دارند.
* این قوانین حداکثر خلاقیت و انعطاف­پذیری را دارا می باشند.
* قوانین فقط برای مجموعه داده که ایجاد شده‌اند دقت خوبی دارد.
* فرایند استخراج قوانین از پرسش‌های کاری خسته­کننده و وقت­گیر می­باشد و نیاز به کار و صرف وقت زیادی است.
* تعداد قوانین استخراج ممکن است فوق العاده زیاد باشد به نحوی که مدیریت آن مشکل گردد.

روش­های مبتنی بر یادگیری، ابتدا نیازمند داده­های برچسب خورده می باشند. سپس مدل­های یادگیری را بر روی این داده­ها طراحی و آموزش می‌دهند. در این دسته روش­ها برای برچسب­گذاری، پرسش‌ها را در قالب مجموعه‌ای از ویژگی­ها در نظر می­گیرند و فرض بر این است که با توجه به ویژگی­های استخراج شده از داده‌های برچسب خورده، الگوها به صورت خودکار استخراج می‌شوند. بنابراین در این روش­ها انتخاب ویژگی­ها و دسته­بندها اهمیت زیادی دارد. ویژگی­های مورد استفاده می­تواند خیلی ساده در حد ویژگی­های کلامی و ریخت­شناسی تا ویژگی­های نحوی و معنایی باشد.

از الگوریتم دسته­بندی مورد استفاده برای دسته­بندی پرسش­ها می­توان به ماشین بردار پشتیبان [84]، شبکه­های بیزی[[105]](#footnote-105) [85]، درخت های تصمیم[[106]](#footnote-106) [85] و SNoW[[107]](#footnote-107) [86] اشاره نمود. البته در میان اینها ماشین بردار پشتیبان بهترین کارایی را داشته است. اما در این بین روش‌های ترکیبی نیز وجود دارد که از هر دو روش مبتنی بر یادگیری و مبتنی بر قاعده استفاده می­کنند [87]، [88].

* **روش­های غیر دسته­بندی**

در این روش دسته­بندی بر روی پرسش انجام نمی­شود و طبیعتاً پرسش­ها برچسب مشخصی نخواهند داشت. روال انجام کار به این صورت است که پرسش­ها به منظور استخراج کلمات کلیدی، مورد پیش پردازش و پردازش قرار می­گیرند. کلمات کلیدی در مرحله بعد با هم ترکیب شده و یک پرس­وجو را می­سازند. مزیت این روش این است که همه پرسش­ها را می­توان به یک صورت پردازش کرد زیرا دسته­های از پیش تعیین شده­ای وجود ندارد. عیب این روش نیز این است که عموماً پرس­وجوهای بدست آمده بی کیفیت هستند و منجر به استخراج پاسخ صحیحی نخواهند شد. درکل، امروزه روش­های مبتنی بر دسته­بندی رایج­ترند [89].

در تدوین پرس­وجو[[108]](#footnote-108)، برای استخراج کلمات کلیدی از تکنیک­های زبانی مانند برچسب‌گذاری[[109]](#footnote-109) [90] و [91]، ریشه­یابی [90]، تجزیه [91] و حذف ایست­واژه­ها [92] و [93] استفاده می­شود. اما مشکل اساسی اینجاست که در غالب اوقات عبارات به کار رفته در پرسش‌ها با عبارات به کار رفته در اسناد حاوی پاسخ های صحیح، یکسان نیستند. این مشکل عدم تطابق[[110]](#footnote-110) نامیده می­شود و یک مساله بسیار مهم و حیاتی در بازیابی اطلاعات می­باشد. این مشکل با استفاده از تکنیک­های تفسیر [96] و [97] و [98] و [99] و بسط [94] و [95] جستجو قابل حل می­باشد. این تکنیک­ها با تولید کلمات یا عبارات جستجوی بیشتر به بازیابی بهتر و مؤثرتر اسناد مرتبط کمک می­کند.

### 2-4-2- بازیابی اسناد

هدف این مرحله این است که با استفاده از یک موتور بازیاب اطلاعات، تعداد کمی از اسناد مرتبط با پرسش مطرح شده، از مجموعه‌ای از اسناد بدون ساختار که احتمالا حاوی پاسخ صحیح به آن پرسش می­باشند، استخراج گردد. این کار به صورت قابل توجهی فضای جستجو را کاهش می­دهد.

در دهه‌های گذشته، مدل‌های بازیابی مختلفی برای بازیابی اسناد توسعه‌یافته­اند، که در میان آنها برخی مدل‌های محبوب مانند: مدل بولی[[111]](#footnote-111)، مدل‌های فضای برداری[[112]](#footnote-112)، مدل‌های احتمالی[[113]](#footnote-113)، مدل‌های زبانی[[114]](#footnote-114) و غیره هستند که به طور خلاصه به شرح زیر مورد بازدید مجدد قرار می‌گیرند.

**مدل بولی**: مدل بولی یکی از ساده‌ترین مدل‌های بازیابی است. پرسش به شکل یک عبارت بولی از اصطلاحات تبدیل می‌شود که با اپراتورهایی مانند "و"، "یا" و "نه" ترکیب می‌شوند، تا دقیقا با اسناد مطابقت داشته باشند و هر سند به عنوان مجموعه‌ای از کلمات دیده شود.

**مدل فضایی بردار**: مدل‌های فضایی بردار، پرسش و هر سند را به عنوان بردارهای کلمه در یک فضای کلمه d بعدی نشان می‌دهند، که در آن d تعداد کلمات در مجموعه واژگان است. هنگام جستجوی اسناد مرتبط برای یک پرسش داده ‌شده، میزان ارتباط هر سند، با محاسبه تشابه (‏برای مثال، شباهت کسینوسی)‏ یا فاصله (‏برای مثال، فاصله اقلیدسی) ‏بین بردار آن سند و بردار پرسش محاسبه می‌شود. در مقایسه با مدل بولی، این رویکرد اسناد مرتبط بیشتری را برای پرسش برمی­گرداند.

**مدل احتمالی**: مدل‌های احتمالی راهی را برای یکپارچه‌سازی روابط احتمالی بین کلمات، در یک مدل، فراهم می‌کنند. برای مثال، پونیاکانوک[[115]](#footnote-115) و همکاران در سال 2008 [101]‏، مدل اوکاپی[[116]](#footnote-116) BM25 که یک مدل احتمالی است را معرفی کردند. این مدل، به فرکانس اصطلاحات و طول اسناد حساس است و یکی از موفق‌ترین مدل‌های بازیابی تجربی است و به طور گسترده در موتورهای جستجوی فعلی استفاده می‌شود.

**مدل زبانی**: مدل‌های زبانی [102]‏ نیز بسیار محبوب هستند، که در میان آن‌ها مدل احتمالی پرس­وجو[[117]](#footnote-117) پیشنهاد شده توسط دیمارنفی[[118]](#footnote-118) و همکاران در سال 2006 [100]‏ بیش‌ترین کاربرد را دارد. در این مدل برای هر سند d یک مدل زبان احتمالی LMd ایجاد می­شود و اسناد براساس احتمال (‏q | LMd‏) P از مدل زبانی که پرسش q را تولید منموده است، رتبه‌بندی می‌شوند.

در عمل، اسناد دریافت ‌شده اغلب حاوی اسناد غیرمرتبط هستند، یا تعداد اسناد آنقدر زیاد است که بیش از ظرفیت مدل استخراج جواب می­باشد. برای پرداختن به این مسائل، پس پردازش در اسناد بازیابی شده بسیار مورد نیاز است. روش‌های متنوعی در پردازش اسناد بازیابی شده که شامل فیلتر کردن اسناد[[119]](#footnote-119)، رتبه‌بندی مجدد اسناد[[120]](#footnote-120) و انتخاب اسناد[[121]](#footnote-121) [7]‏ و غیره است، استفاده شده است. از فیلتر کردن اسناد برای شناسایی و حذف نویز با توجه به پرسش ورودی استفاده می­شود؛ رتبه‌بندی مجدد اسناد، به منظور مرتب‌سازی بیشتر اسناد، با توجه به درجه قابل قبول بودن اسناد حاوی پاسخ، به ترتیب نزولی توسعه داده می‌شود؛ انتخاب سند، انتخاب اسناد و مدارک مرتبط با آن است. پس از پس­پردازش، تنها اسناد مرتبط باقی مانده و در اختیار مرحله بعد برای استخراج پاسخ نهایی قرار می­گیرند.

### 2-4-3- استخراج پاسخ

هدف نهایی یک سیستم پرسش و پاسخ، پاسخ موفقیت‌آمیز به پرسش‌های داده ‌شده بوده و مرحله استخراج پاسخ، مسئول بازگرداندن دقیق‌ترین پاسخ به یک پرسش کاربر می‌باشد. عملکرد این مرحله با پیچیدگی پرسش، انواع پاسخ مورد انتظار از مرحله تجزیه و تحلیل پرسش، اسناد بازیابی شده از مرحله بازیابی اسناد و همچنین روش استخراج پاسخ اتخاذ شده و غیره درگیر می­باشد. باتوجه به اهمیت این مرحله، پژوهشگران می­بایست توجه ویژه‌ای به این مرحله داشته باشند و جایگاه ویژه‌ای برای آن قائل باشند.

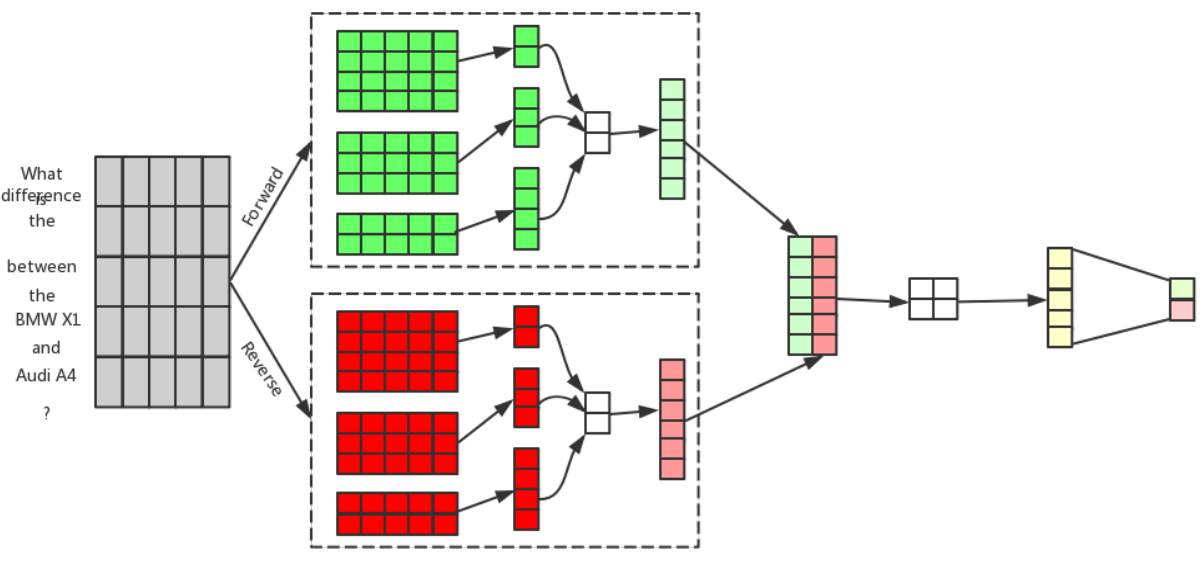
در سیستم‌های سنتی پرسش و پاسخ، پرسش­های ساده و پرسش­های لیستی [103]‏ به طور گسترده‌ و برای مدتی طولانی مورد مطالعه قرار گرفته‌اند. به طور معمول، پاسخ‌ها به پرسش‌های حقیقی[[122]](#footnote-122) (‏برای مثال، کی، کجا، چه کسی …)،‏ یک متن ساده در اسناد از قبیل: یک نام واحد، یک نشانه کلمه یا یک عبارت اسم می­باشد. در حالی که پرسش­های لیستی پرسش­هایی هستند که پاسخ‌های آنها مجموعه‌ای از عوامل هستند که از یک سند یکسان یا از اسناد مختلف جمع‌آوری شده‌اند. نوع پاسخ دریافت‌شده از مرحله تجزیه و تحلیل پرسش نقش مهمی ایفا می‌کند، به خصوص برای پرسش­هایی که پاسخ‌های آنها موجودیت‌های نام­گذاری شده[[123]](#footnote-123) هستند. بنابراین، سیستم‌های اولیه به شدت بر تکنیک شناسایی نام گذاری شده (‏NER) ‏[[124]](#footnote-124)تکیه می‌کنند [9۰]، [104]، [105]‏، زیرا مقایسه موجودیت­های شناخته‌شده و نوع پاسخ ممکن است به راحتی پاسخ نهایی را نتیجه دهد. ویلسون[[125]](#footnote-125) و همکاران در سال 1994 [106]‏، پیشنهاد کردند که استخراج پاسخ به عنوان یک فرآیند واحد توصیف شود. روش پیشنهادی آنها به این صورت است که ابتدا اطلاعات پنهان، به ترتیب از پرسش و پاسخ کشف شده، و سپس از برخی روش‌های تطبیقی مانند انطباق الگوی متن سطحی [107]‏، [‏108]‏، انطباق کلمات یا عبارات [71]‏، و انطباق ساختار نحوی [‏9۰]‏، [93]‏، [109]‏ برای تشخیص پاسخ‌ها، استفاده شود.

می­بایست از صحت پاسخ استخراج‌شده پیش از ارائه به کاربران نهایی اطمینان حاصل کرد. به علاوه، در برخی موارد ممکن است چندین پاسخ برای یک پرسش تولید شود و باید یکی از آنها انتخاب شود. برای حل چنین مسائلی از روش اعتبارسنجی جواب استفاده شده ‌است. یکی از روش‌های اعتبارسنجی که به طور گسترده مورد استفاده قرار می‌گیرد، استفاده از یک منبع اطلاعاتی اضافی مانند یک موتور جستجوی وب برای اعتبارسنجی صحت هر پاسخ کاندید است. اساس بر این است که سیستم باید تعداد زیادی از اسناد که حاوی عبارات پرسش و پاسخ هستند را برگرداند. هر چه تعداد این اسناد بازگشتی بیشتر باشد، احتمال اینکه پاسخ صحیح باشد بیشتر خواهد بود [7].

## **2-5- کاربرد شبکه‌های عصبی عمیق در سیستم­های پرسش و پاسخ**

در دهه اخیر، تکنیک‌های یادگیری عمیق نیز با موفقیت به سیستم­های پرسش و پاسخ اعمال شده‌اند. یادگیری عمیق تقریباً در هر مرحله از یک سیستم پرسش و پاسخ مورد استفاده قرار می­گیرد و علاوه بر آن، این امکان را به سیستم­های پرسش و پاسخ می‌دهد که به صورت پایانه به پایانه آموزش­پذیر باشند.

در فاز تحلیل پرسش، برخی مقالات، طبقه‌بندی کننده‌های عصبی را برای تعیین انواع پرسش­ها پیشنهاد می­کنند. به عنوان مثال، لی و همکاران در سال 2018 [110]‏ از یک مدل مبتنی بر CNN به نام CNN معکوس برای طبقه­بندی پرسش استفاده نمودند. در این روش یک مکانیزم معکوس به معماری CNN افزوده می­شود. این معماری در شکل 2-12 قابل مشاهده است. کارایی این روش با استفاده از معیار دقت اندازه گیری شده است و این معیار، در مجموعه داده­های TREC و MPQA به ترتیب مقادیر 8/95 و 2/89 می­باشد.



شکل 2-12- معماری CNN معکوس برای طبقه­بندی پرسش [110]

همچنین ضیا[[126]](#footnote-126) و همکاران در سال 2018 [111]‏، از یک مدل مبتنی بر LSTM به نام Att-LSTM سلسله مراتبی برای طبقه­بندی پرسش استفاده نمودند. این مدل دارای 5 ماژول است که عبارتند از:

الف) جاسازی حروف: پرسش به حروف شکسته شده و جاسازی می­گردد.

ب) Att-LSTM سطح پایین: بازنمایی لغات را از جاسازی حروف بدست آمده از ماژول قبل یاد می­گیرد.

ج) شبکه Highway: بازنمایی لغات بدست آمده از ماژول قبل را بهینه می­کند و از این طریق اطلاعات بهینه­تری را برای ماژول بعد مهیّا می­سازد.

د) ماژول Att-LSTM سطح بالا: بازنمایی­های لغات بهینه شده را به عنوان ورودی گرفته و بازنمایی جملات را یاد می­گیرد.

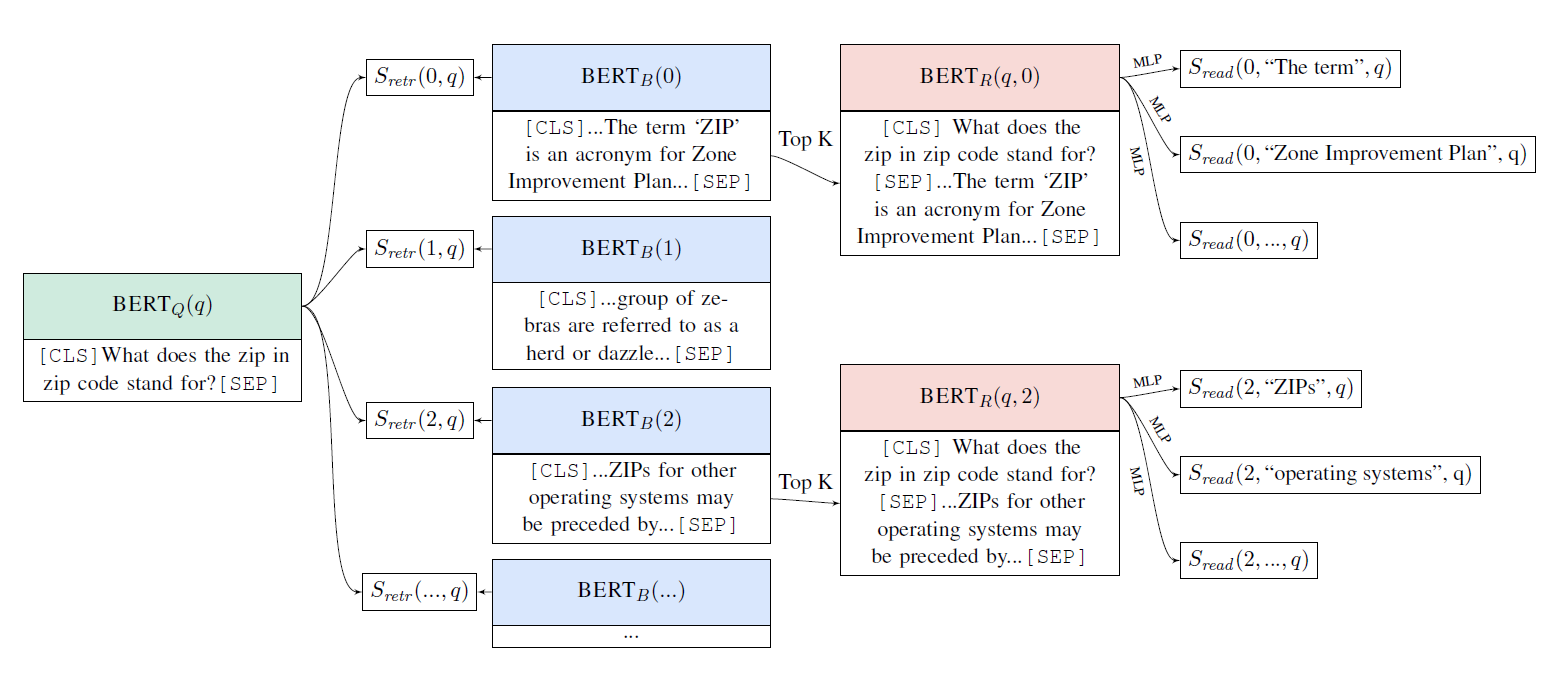
ه) لایه Softmax: در نهایت بازنمایی جمله به عنوان ورودی به لایه Softmax داده می­شود. در این لایه احتمال تعلق پرسش ورودی به یکی از کلاس­های تعیین شده محاسبه می­گردد.

در این روش معیار کارایی دقت بر روی مجموعه داده­های MNIST و pMNIST به ترتیب مقادیر 9/98 و 2/96 گزارش شده است.

در بحث بازیابی اسناد، برای رفع مساله "عدم تطابق"، از روش­های مبتنی بر نمایش متراکم استفاده می­شود. در روش­های بازیابی عمیق، بر خلاف روش‌های سنتی مانند TF - IDF و BM25 که از بازنمایی­های پراکنده[[127]](#footnote-127) استفاده می‌کنند، از روش‌های مبتنی بر بازنمایی متراکم[[128]](#footnote-128) [‏۱12]‏، [113]‏، [114]‏، [11۵]‏ استفاده می­شود. روش‌های بازیابی عمیق یاد می‌گیرند که پرسش­ها و اسناد را به یک فضای برداری نهفته کدگذاری کنند که در آن معناشناسی متن فراتر از تطابق اصطلاحات در نظر گرفته می­شود. برای مثال، داس[[129]](#footnote-129) و همکاران در سال 2019 [113]‏ در سیستم پرسش و پاسخ پیشنهادی خود از سه بخش بازیاب پاراگراف، خواننده و یک multi-step-rasoner استفاده کردند. در بخش بازیاب پاراگراف، امتیاز ارتباط هر پاراگراف و پرسش ورودی محاسبه و رتبه­بندی می­شود. در بخش خواننده k پاراگراف با امتیاز بالاتر را به عنوان ورودی گرفته و در صورت امکان محدوده پاسخ را در آنها مشخص می­نماید. در نهایت در بخش multi-step-rasoner که یک GRU[[130]](#footnote-130) می­باشد. بر اساس پرس­وجو و حالت فعلی خواننده به صورت مکرر پرس و جو را به روز کرده و آن را در اختیار خواننده و بازیاب قرار می­دهد تا پاراگراف­ها و محدوده­های پاسخ به روز رسانی شوند. معیارهای کارایی EM و F1 در مجموعه داده TRIVIAQA-unfiltered به ترتیب مقادیر 56/61 و 03/68 و در مجموعه داده SearchQA مقادیر 36/61 و 26/56 را کسب کرده است. همچنین فلدمن[[131]](#footnote-131) در سال 2019 [11۵]‏ در مدل پیشنهادی خود، کدکننده‌های خودش را آموزش می‌دهد و هر سند و پرسش را به صورت مستقل در بردارهای متراکم کدگذاری کرده، و امتیاز شباهت بین آنها را با استفاده از ضرب داخلی بردارهای آنها (MIPS) محاسبه می­نماید. الگوریتم MIPS[[132]](#footnote-132) ‏[116]‏، [117]‏، [118]‏ برای بهبود کارایی بازیابی، هنگامی که مخزن اسناد، بزرگ مقیاس باشد، مورد استفاده قرار می‌گیرد. همچنین در این روش از مدل‌های عصبی به عنوان یک مرحله قطعی در استخراج پاسخ، استفاده می­شود. استخراج پاسخ‌ها از برخی اسناد مرتبط به یک پرسش، اساساً همان کار درک مطلب ماشینی (‏MRC) را انجام می‌دهد. در این روش معیارهای کارایی EM و F1 در مجموعه­ داده SQuAD-Openمورد بررسی قرار گرفته که به ترتیب 2/46 و 3/39 می­باشد.

در چند سال گذشته، با ظهور برخی از مجموعه داده‌های بزرگ مقیاس مانند CNN / Daily Mail [۱19]‏، MS MARCO ‏ [120]‏، RACE‏ [1۲۱]‏ و SQuAD 0.2 [‏1۲۲]‏، تحقیقات در مورد درک مطلب ماشین­های عصبی به پیشرفت قابل‌توجهی دست یافته‌ است [37]‏، [124]‏، [125]‏، [40]‏. به عنوان مثال، سئو[[133]](#footnote-133) در سال 2017 [1۲3]‏ با پیشنهاد مدل BiDAF، اسناد بازیابی شده را از طریق ساختار سلسله مراتبی چند مرحله‌ای، در سطوح (لایه­های) مختلف که شامل یک لایه نگاشت کاراکتر، یک لایه نگاشت کلمه، و یک لایه نگاشت متن است، نمایش می‌دهد، و یک مکانیزم جریان توجه دو طرفه را برای به دست آوردن بازنمایی سند-پرسش به کار می­گیرد. در این روش معیار کارایی EM در مجموعه داده CNN مقدار 9/76 و در مجموعه داده DailyMail 6/79 گزارش شده است. همچنین دوهان[[134]](#footnote-134) و همکاران در سال 2018‏ [125]‏ در مدل پیشنهادی خود تحت عنوان QANet ، از CNN و مکانیزم خود-توجهی[[135]](#footnote-135) [‏127]‏ و همچنین تعاملات محلی و سراسری برای مدل‌سازی استفاده می‌کنند، که به طور قابل ‌توجهی سریع‌تر از مدل‌های تکرار شونده معمول عمل می‌کند. این روش دو نقطه ابتدا و انتهای پاسخ، در یک متن را به عنوان خروجی بر می­گرداند و و ارزیابی آن بر روی مجموعه داده SQuAD مقدار 1/51 را برای معیار EM و 6/56 را برای معیار F1 گزارش کرده است.

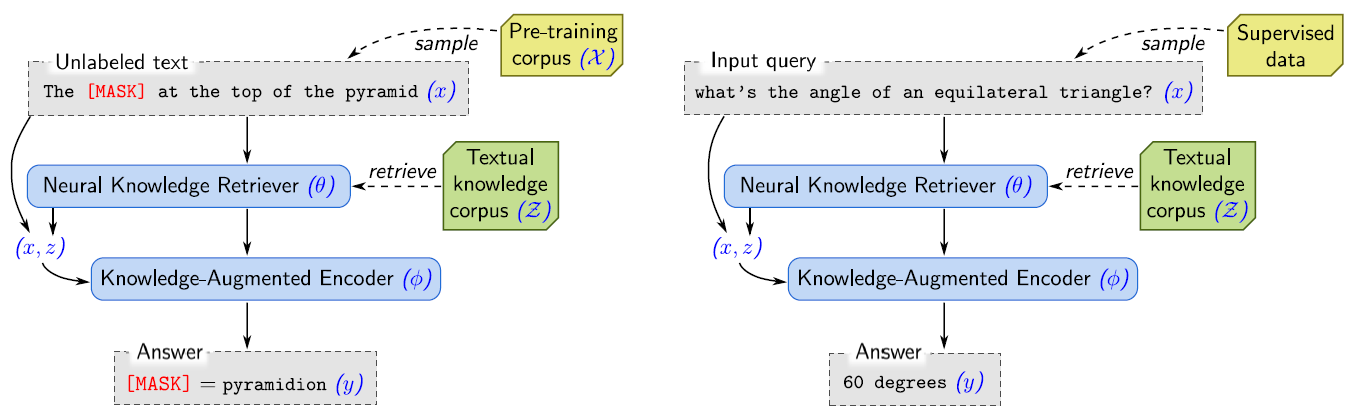
علاوه بر این، استفاده از یادگیری عمیق، سیستم‌های پرسش و پاسخ را قادر می‌سازد تا به صورت پایانه به پایانه قابل آموزش باشند [‏28]، [44]، [49]‏. به عنوان مثال، لی[[136]](#footnote-136) و همکاران در سال 2019 [49]‏ مدلی تحت عنوان ORQA پیشنهاد کردند که در آن مرحله بازیابی اسناد، به عنوان متغیر پنهان در نظر گرفته شده در سیستم پرسش و پاسخ ادغام می­گردد. در این حالت، سیستم تنها بر اساس جفت رشته­های پرسش-پاسخ و با استفاده از شبکه برت [40] آموزش می­بیند. شمای کلی این مدل در شکل 2-13 قابل مشاهد است.



شکل 2-13- شمای کلی از مدل ORQA [49]

در این مدل، زیر مجموعه‌ای از تمام مشتقات ممکن پاسخ، برای پرسش q در شکل 2-13 نشان‌داده شده ‌است. امتیازات بازیابی (‏b;q)Sretr ‏از طریق ضرب داخلی بین کدکننده‌های مبتنی بر برت محاسبه می‌شوند. بلوک­هایی که امتیاز بالاتری بدست می­آورند مشترکاً با پرسش کدگذاری می­شوند. سپس بازه پاسخ Sread (b; s; q) با استفاده از یک پرسپترون چند لایه امتیازدهی می­شود. امتیاز نهایی مدل نیز از مجموع امتیاز بازیابی و پاسخ محاسبه می­گردد. معیار EM در این روش بر روی دو مجموعه داده TriviaQA و SQuAD به ترتیب 1/47 و 2/33 گزارش شده است.

گو[[137]](#footnote-137) و همکاران در سال 2020 [44]‏ مدلی تحت عنوان REALM پیشنهاد کردند که یک مدل زبانی پیش­آموزشی و شامل یک بازیاب و یک کدگذار دانش است. در این مدل، بازیاب دانش و کدگذار، شبکه‌های عصبی متمایز هستند، که قادر به محاسبه گرادیان با توجه به پارامترهای مدل می­باشند که در سراسر شبکه منتشر می‌شود. این مدل، شبیه به دیگر مدل‌های زبانی پیش آموزشی، دارای دو مرحله پیش آموزش و تنظیم پارامترها است. در مرحله پیش آموزش، مدل به شیوه‌ای بدون نظارت، با استفاده از مدل‌سازی زبان پوششی[[138]](#footnote-138) آموزش داده می‌شود، در حالی که پارامترها با استفاده از مثال‌های نظارت شده در مرحله تنظیم پارامترها، تنظیم می‌شوند. شمای کلی این مدل در شکل 2-13 قابل مشاهد است.



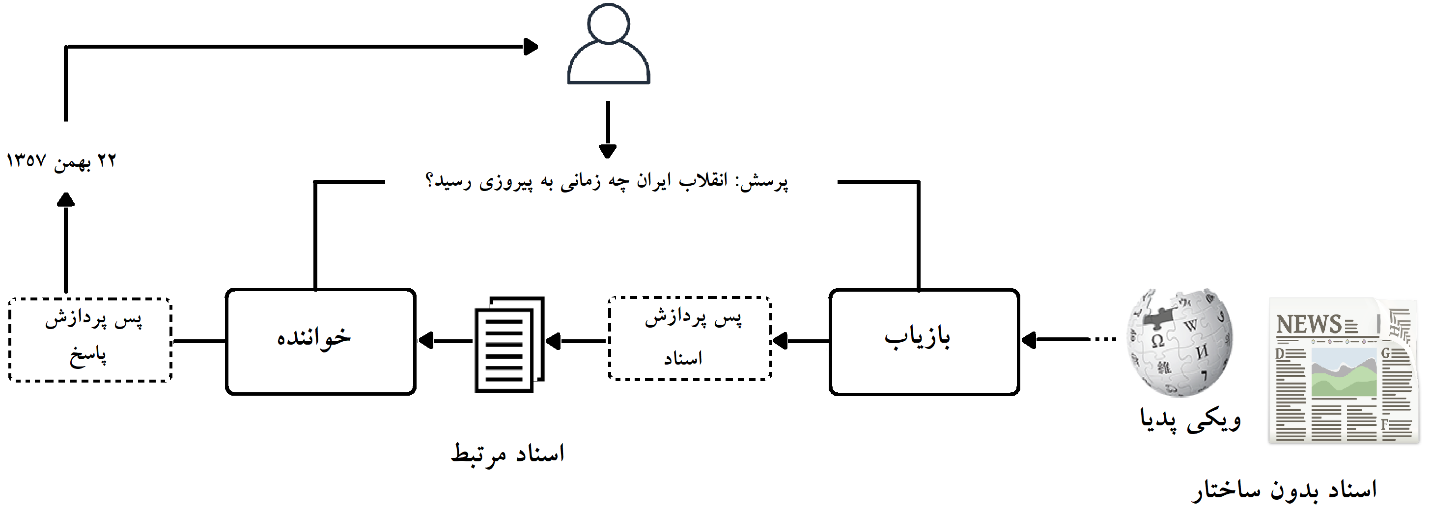
شکل 2-14- شمای کلی مدل REALM [44]

همان­گونه که در سمت چپ شکل 2-14 قابل مشاهده است، بازیاب و کدگذار دانش به صورت بدون نظارت پیش آموزش می­شوند و در سمت راست مقادیر پارامترها تنظیم می­شوند. این روش بر روی سه مجموعه داده NaturalQuestions-Open، WebQuestions و CuratedTrec مورد ارزیابی قرار گرفته و معیار EM برای هریک به ترتیب 4/40، 7/40 و 8/46 گزارش شده است.

در سیستم‌های پرسش و پاسخ اولیه، موفقیت در پاسخگویی به شدت وابسته به عملکرد تحلیل پرسش، به ویژه طبقه‌بندی پرسش است که انواع پاسخ‌های مورد انتظار را فراهم می‌کند [‏70]‏ که توسط زبان­شناسان ساخته می­شوند. اما این طبقه­بندی­ها، غیر بهینه هستند زیرا پوشش دادن تمام انواع پرسش­ها در واقعیت، بسیار پیچیده و غیر ممکن است. علاوه بر این، خطاهای طبقه‌بندی به راحتی منجر به شکست استخراج پاسخ می‌شوند، در نتیجه به شدت به عملکرد کلی سیستم آسیب می‌زنند. براساس آزمایش‌های انجام‌شده، حدود ۳۶.۴ % از خطاها در سیستم‌های پرسش و پاسخ اولیه، ناشی از عدم طبقه‌بندی صحیح انواع پرسش‌ها می‌باشد [70]‏. مدل‌های عصبی قادر هستند تا پرسش­های به زبان طبیعی را به بازنمایی­هایی که برای ماشین‌ها قابل‌تشخیص هستند، تبدیل نمایند. علاوه بر این، مدل‌های درک مطلب ماشین عصبی یک راه‌حل قدرتمند برای استخراج پاسخ در سیستم­های پرسش و پاسخ ارائه می‌دهند، که تا حد زیادی ضرورت استفاده از تکنیک‌های تحلیلی زبانی سنتی برای پرسش­ها را مرتفع ساخته و انقلابی برای سیستم‌های پرسش و پاسخ به ارمغان می­آورند [‏41]، [42]، [43]، [49]‏. نخستین کار در این حوزه توسط چن[[139]](#footnote-139) و همکاران در سال 2017 [41] تحت عنوان سیستم DrQA ارائه شده است که مدل‌های درک مطلب ماشین عصبی را با سیستم­های پرسش و پاسخ ترکیب می­کند. در واقع در این کار معماری "بازیاب - خواننده" توضیح داده شده در قبل، تکامل می‌یابد. این روش، تکنیک بازیابی اطلاعات مبتنی بر TF-IDF و مدل درک مطلب ماشین عصبی را با هم ترکیب کرده و نتایج قابل قبولی در حوزه پاسخ به پرسش­های حقیقی دامنه باز دریافت کرده است. در واقع خروجی خواننده در روش DrQA، شامل محل قرار گرفتن نشانه آغازین و پایانی پاسخ در یک پاراگراف می­باشد و ارزیابی آن بر روی مجموعه داده SQuAD مقدار 70 درصد برای معیار EM و 79 درصد برای معیار F1 گزارش شده است. پس از DrQA‏ کارهای زیادی در این حوزه منتشر شده ‌است [42]‏، [44]‏، [47]‏، [48]‏، [49]‏، [37]‏، [128]‏، [129]‏.

## **2-6- سیستم­های پرسش و پاسخ مدرن**

در این بخش، همان طور که در شکل 2-15 نشان‌داده شده ‌است، معماری "بازیاب - خواننده" در سیستم­های پرسش و پاسخ را معرفی می‌کنیم. معماری سیستم­های پرسش و پاسخ مدرن از دو مولفه اصلی تشکیل شده است. 1) بازیاب: اسناد مرتبط با یک پرسش مشخص را بازیابی می‌کند، و می‌تواند به عنوان یک سیستم بازیابی اطلاعات در نظر گرفته شود، 2) خواننده: پاسخ نهایی را از اسناد دریافتی استنتاج می­نماید، و معمولاً یک مدل درک مطلب ماشین عصبی است. به علاوه برخی ماژول­های کمکی دیگر که در شکل 2-15 با خط تیره مشخص شده است می­تواند در یک سیستم پرسش و پاسخ گنجانده شود. این ماژول­ها شامل پس­پردازش اسناد[[140]](#footnote-140)، که اسناد بازیابی شده را به منظور یافتن مرتبط­ترین آنها، فیلتر، و مجدداً رتبه­بندی می­کند. پس­پردازش پاسخ[[141]](#footnote-141)، که پاسخ نهایی را از میان پاسخ­های کاندید انتخاب می­کند. سیستم­هایی که از این معماری تبعیت می­کنند به دو دسته سیستم­های خط لوله[[142]](#footnote-142) و سیستم­های پایانه به پایانه[[143]](#footnote-143) تقسیم می­شوند. شکل 2-16 یک طبقه­بندی از سیستم­های پرسش و پاسخ



شکل 2-15- نمونه­ای از معماری "بازیاب – خواننده" در سیستم­های پرسش و پاسخ [193]

مدرن را نشان می­دهد که به درک بهتر آن کمک می­کند. در ادامه ابتدا اجزاء سیستم­های خط لوله را بررسی و در ادامه سیستم­های پرسش و پاسخ قابل آموزش پایانه به پایانه دنبال خواهد شد.

### 2-6-1- سیستم­های مبتنی بر خط لوله در سیستم­های پرسش و پاسخ مدرن

همانگونه که در شکل 2-15 قابل مشاهده است، سیستم­های مبتنی بر خط لوله در سیستم­های پرسش و پاسخ مدرن از تعدادی اجزا تشکیل شده که با یکدیگر در تعامل هستند و در ادامه هر یک از اجزا بررسی خواهند شد.

### 2-6-1-1- بازیاب

بازیاب معمولاً به عنوان یک سیستم بازیابی اطلاعات در نظر گرفته می‌شود و هدف آن، بازیابی اسناد یا متون مرتبط حاوی پاسخ است. به طور کلی، رویکردهای فعلی برای بازیابی را می­توان به سه دسته تقسیم کرد، که عبارتند از: بازیاب تنک[[144]](#footnote-144)، بازیاب متراکم[[145]](#footnote-145)، و بازیاب تکرارگر[[146]](#footnote-146)، که در ادامه به تفصیل شرح داده خواهند شد.

بازیاب تنک

Retrieve-and-Read, ORQA, REALM, RAG

DS-QA, InferSent Re-ranker, Relation-Networks Reranker,

Paragraph Ranker,

DenSPI

BART Reader, T5 Reader

Multi-Passage BERT RerankerMulti-Passage BERT Reranker

DenSPI, ORQA, REALM, DPR, ColBERT, SPARTA

TF-IDF, BM25, DrQA,

BERTserini

DrQA, Match-LSTM, BiDAF, S-Norm Reader, BERT

Reader, Graph Reader

Strength-based Re-Ranker

Golden Retriever, Multistep Reasoner, Adaptive Retriever, Path Retriever, MUPPET, DDRQA, MDR, Graph Retriever, GAR

Coverage-Based Re-Ranker, RankQA

بازیاب

بازیاب متراکم

بازیاب تکرارگر

آموزش نظارتی

پس­پردازش اسنادناد

آموزش بدون نظارت

سیستم­های پرسش و پاسخ مدرن

انتقال آموزش

خواننده استخراج­گر ااستخراج­گر

خواننده

خواننده مولد

پس­پردازش پاسخ

مبتنی بر قانون

مبتنی بر آموزش

بازیاب- خواننده

پایانه به پایانه

بازیاب - تنها

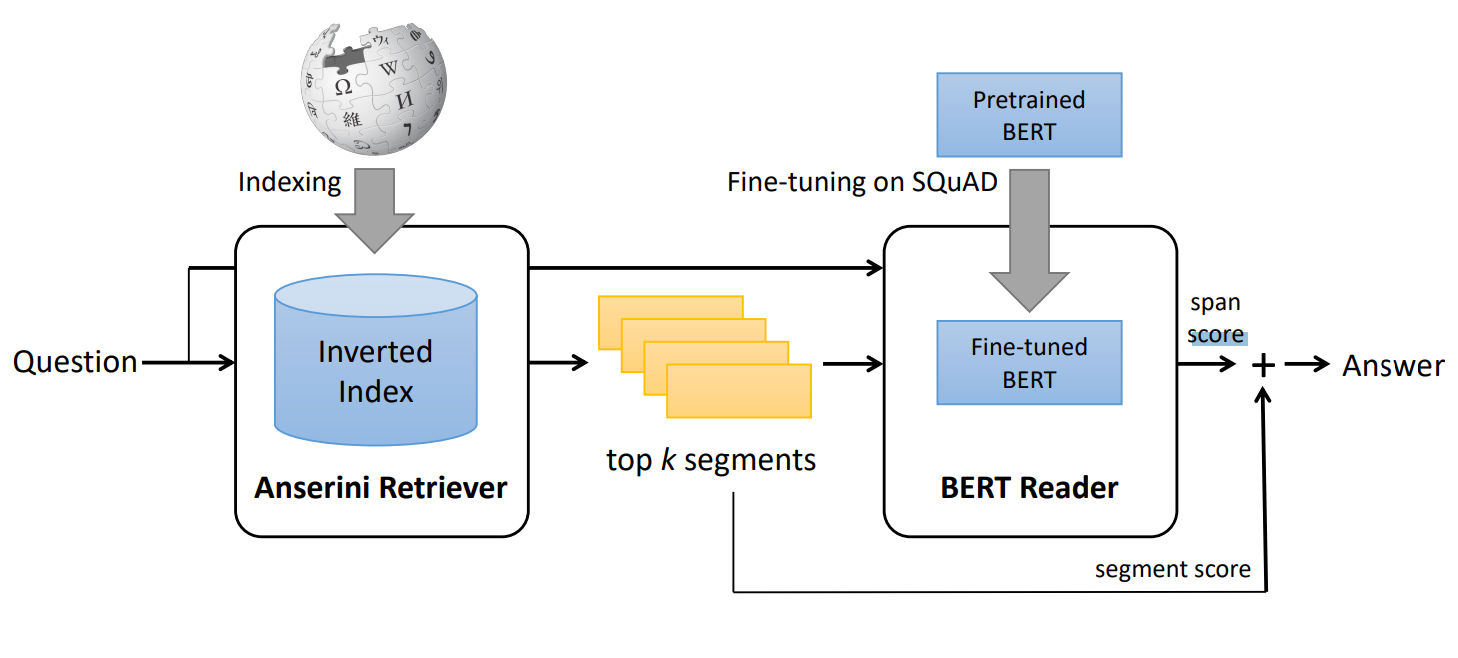
GPT2, GPT3, T5, BART

بازیاب - آزاد

شکل 2-16- طبقه­بندی در سیستم­های پرسش و پاسخ مدرن

### 2-6-1-1-1- بازیاب تنک

این روش به سیستم‌هایی اشاره می‌کند که با اتخاذ روش‌های بازیابی کلاسیک که پیش­تر معرفی شده‌اند، به دنبال یافتن اسناد مرتبط هستند، مانند TF - IDF ‏ [41]‏، [48]‏، [1]‏، [131]‏ و BM25 [132]‏، [133]‏. مدل DrQA پیشنهاد شده توسط چن و همکاران [‏41] که پیش­تر معرفی شد، ‏اولین رویکرد برای سیستم‌های پرسش و پاسخ مدرن است که در آن تکنیک‌های بازیابی کلاسیک مورد استفاده قرار می­گیرد. بازیاب در DrQA، از هشینگ بایگرام [‏134]‏ و تطابق TF - IDF برای جستجو بر روی ویکی‌پدیا استفاده می‌کند. مدل دیگری که از بازیابی تنک استفاده می­نماید توسط یانگ و همکارانش در سال 2019 [132] به نام BERTserini توسعه داده شد که در آن از بازیابی Anserini  [135] ‏استفاده شده است. Anserini، یک ‌ابزار بازیابی اطلاعات متن باز مبتنی بر Lucene  می­باشد.



2-17- معماری BERTserini [132]

همانگونه که در شکل 2-17 قابل مشاهده است، معماری این مدل از دو بخش خواننده (خواننده برت) و بازیاب (بازیاب Anserini) تشکیل شده است. در این روش، جزئیات متن در سطوح مختلف، از جمله سطح مستند، سطح پاراگراف و سطح جمله به صورت تجربی بررسی می­شوند و نتایج نشان می‌دهد که جزئیات بدست آمده از سطح پاراگراف بهترین کارایی را دارد. روال کار به این صورت است که K قطعه برتر مرتبط با پرسش بدست آمده از بازیاب به همراه امتیاز قطعات (امتیاز قطعه نشان دهنده میزان احتمال حاوی جواب بودن قطعه است و با SAnserini نشان داده می­شود)، به عنوان ورودی به خواننده داده می­شود. خواننده برت که مبتنی بر مدل برت پایه است به تمامی قطعات اعمال شده و محدوده پاسخ به همراه امتیاز آن محدوده (امتیاز محدوده، امتیازی است که خواننده برای یک محدوده از یک قطعه به عنوان پاسخ پرسش در نظر گرفته است.) برگردانده می­شود. در نهایت پاسخ از رابطه حاصل می­شود. در این رابطه μ یک هایپرپارامتر است که مقداری بین صفر و یک دارد. این روش بر روی مجموعه داده SQuAD به مقدار 6/38 برای معیار EM و 1/46 برای معیار F1 دست یافته است.

روش­های بازیابی سنتی مانند TF-IDF و BM25 از بازنمایی­های تنک، برای تعیین میزان تطابق اصطلاحات استفاده می‌کنند. اما، عبارات مورد استفاده در پرسش­های کاربر اغلب با عبارات موجود در اسناد یکسان نیستند. روش‌های مختلفی براساس بازنمایی‌های متراکم [112]‏، [113]‏، [44]‏، [115]‏ در سال‌های اخیر توسعه‌یافته­اند، که در آنها پرسش­ها و اسناد به یک فضای برداری نهفته نگاشت می­شوند که در این فضای برداری، معناشناسی متن فراتر از تطابق واژه در نظر گرفته می­شود.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| الف) بازیاب مبتنی بر بازنمایی | ب) بازیاب مبتنی بر تقابل | ج) بازیاب مبتنی بر تقابل-بازنمایی |

شکل 2-18- سه نوع بازیاب متراکم [28]، [30]، [112]

### 2-6-1-1-2- بازیاب متراکم

در کنار موفقیتی که یادگیری عمیق با پیشنهاد بازنمایی معنایی داشت، مدل‌های بازیابی عمیق مختلفی در چند سال گذشته توسعه داده شده‌اند که اثربخشی بازیابی و در نتیجه آن، عملکرد نهایی سیستم­های پرسش و پاسخ را به شدت افزایش داده­اند. همان­طور که در شکل2-18 نشان‌ داده شده ‌است، بازیاب متراکم در سیستم‌های پرسش و پاسخ با توجه به روش‌های مختلف رمزگذاری پرسش و سند و همچنین امتیاز شباهت آنها، به سه دسته تقسیم می­شوند: بازیاب مبتنی بر بازنمایی [‏112]‏، [44]‏، [49]‏، [136]‏، بازیاب مبتنی بر تعامل [‏28]‏، [46]‏ و بازیابی مبتنی بر بازنمایی- تعامل [30]‏، [82]‏، [138]‏.

**الف) بازیاب مبتنی بر بازنمایی:** بازیابی مبتنی بر بازنمایی، که به آن بازیاب کدگذار دوگانه[[147]](#footnote-147) یا بازیاب دو برج[[148]](#footnote-148) گفته می‌شود، از دو رمزگذار مستقل مانند برت [‏40]‏ برای کدگذاری پرسش و سند استفاده می‌کند و میزان ارتباط آنها را با محاسبه امتیاز تشابه بین دو بازنمایی، برآورد می‌کند. از جمله سیستم­هایی که از بازیاب مبتنی بر بازنمایی استفاده می­کنند می­توان به ORQA [49]‏ (توضیح داده شده در بخش 2-5) که دو کدگذار مستقل مبتنی بر برت را برای کدگذاری پرسش و سند اتخاذ می‌کند و امتیاز ارتباط بین آنها را توسط ضرب داخلی بردارهای آنها محاسبه می‌کند، اشاره نمود. همچنین کارپوخین[[149]](#footnote-149) و همکاران در سال 2020 [112] مدلی تحت عنوان DPR پیشنهاد کردند که‏ همانند ORQA از دو رمزگذار برت مستقل استفاده می‌کند، اما مرحله پیش آموزش را به دلیل پر هزینه بودن آن حذف می‌کند و به جای آن، بر یادگیری یک بازیاب قوی با استفاده از جفت­های پرسش و پاسخ‌ متمرکز است. این روش بر روی مجموعه داده­های NQ، TriviaQA، WQ، TREC و SQuAD مورد بررسی قرار گرفته و برای معیار EM به ترتیب مقادیر 5/41، 8/56، 4/42، 4/49 و 1/24 را گزارش کرده است. روش مبتنی بر بازنمایی [112]‏، [44]‏، [49]‏ می‌تواند بسیار سریع باشد زیرا بازنمایی­های اسناد را می­توان از پیش محاسبه و به صورت آفلاین نمایه نمود. اما در این حالت، ممکن است کارایی افت کند، زیرا از آنجایی که بازنمایی پرسش و سند به طور مستقل به دست می‌آیند، در نتیجه تنها منجر به تعاملات سطحی بین آن‌ها می­گردد.

**ب) بازیابی مبتنی بر تعامل:** چنین بازیاب­هایی یک پرسش و اسناد مرتبط با آن را به عنوان ورودی می‌گیرند، و آنها را بر اساس تعاملات نشانه­ها در سطوح مختلف، مدل­سازی می­کنند، که منجر به خلق یک بازیاب قدرتمند می­شود [40]‏، [127]‏.

نیشیدا[[150]](#footnote-150) و همکاران در سال 2018 [28]‏، از بازیابی مبتنی بر تعامل بهره برده و پیشنهاد آموزش مشترک بازیاب و خواننده، با استفاده از یادگیری چند وظیفه نظارتی را می‌دهد‏. روش پیشنهادی آنها از شش لایه تشکیل شده است که عبارتند از:

* لایه جاسازی لغات، که وظیفه نگاشت لغات به فضای برداری را دارد.
* لایه جاسازی زمینه، که وظیفه کدگذاری روابط بین لغات را دارد.
* لایه توجه، که وظیفه جفت کردن پرسش و متون و در نهایت تولید بردار لغات پرسش-متن را به عهده دارد.
* لایه مدل­سازی، وظیفه جمع­آوری ارتباط لغات متون با توجه به پرسش را به عهده دارد.
* لایه ادراک، که یک پاسخ به پرسش ورودی را نمایش می­دهد.
* لایه بازیاب، که متون مرتبط به پرسش را برمی­گرداند.

این روش بر روی مجموعه داده­ SQuAD مورد بررسی قرار گرفته و برای معیار EM به ترتیب مقدار 6/35 و برای معیار F1 مقدار 6/42 را گزارش کرده است.

سئو و همکاران در سال 2017 [37] با استفاده از بازیابی مبتنی بر تعامل، روشی بر اساس BiDAF‏، پیشنهاد کردند که در آن یک لایه بازیابی برای محاسبه امتیاز ارتباط بین پرسش و سند اضافه می‌شود. به علاوه یک لایه درک[[151]](#footnote-151) برای پیش‌بینی موقعیت آغاز و پایان محدوده پاسخ نیز در نظر گرفته می‌شود [28]. این روش در مجموعه داده SQuAD برای معیارهای EM و F1 به ترتیب مقادیر 3/73 و 1/81 را کسب کرده است.

از دیگر روش­های بازیابی مبتنی بر تعامل می­توان به مدل ارائه شده توسط نی[[152]](#footnote-152) و همکاران در سال 2019 [46]‏ که یک بازیاب متراکم در سطح پاراگراف و مدل ارائه شده توسط دِلوین و همکاران در سال 2018 [40]‏ که یک بازیاب متراکم در سطح جمله را براساس برت توسعه می‌دهند. بازیاب­های مذکور، فرآیند بازیابی متراکم را به عنوان یک مساله طبقه‌بندی دودویی در نظر می‌گیرند و جفت پرسش و سند را به عنوان ورودی گرفته و از نگاشت نشانه [CLS] برای تعیین اینکه آیا مرتبط هستند یا نه، استفاده می‌کنند. آزمایش‌ها نشان می‌دهد که هر دو بازیاب چه در سطح پاراگراف و چه در سطح جمله برای به دست آوردن عملکرد خوب سیستم ضروری هستند. این روش بسیار قدرتمند است، زیرا امکان تعاملات بسیار قوی بین پرسش و سند را فراهم می‌آورد. در هر صورت، چنین روشی معمولا نیاز به محاسبات سنگینی دارد که گاهی بسیار پرهزینه است که کاربرد آن را برای اسناد بزرگ دشوار می‌سازد. معیار کارایی F1 در روش ارائه شده توسط نی [46] در مجموعه داده HOTPOTQA مقدار 2/76 را کسب کرده است. همچنین معیار کارایی F1 در مدل ارائه شده توسط دلوین [40] در مجموعه داده SQuAD مقدار 1/83 را کسب کرده است.

**ج) بازیاب تعامل- بازنمایی:** به منظور دستیابی به دقت و کارایی بالا، اخیراً برخی از سیستم‌های روش‌های مبتنی بر بازنمایی و مبتنی بر تعامل را ترکیب می‌کنند [30]، [82]، [138]. به عنوان مثال، خَتاب و همکاران در سال 2020 [30]، با استفاده از یک بازیاب مبتنی بر تعامل-بازنمایی، یک سیستم پرسش و پاسخ مبتنی ColBERT-QAرا پیشنهاد کردند که در این سیستم، قسمت بازیاب براساس ColberT [139]‏ توسعه می­یابد. در این توسعه به منظور محاسبه امتیاز شباهت، از معماری کدگذار دوگانه بر روی بازنمایی­های پرسش و سند استفاده می­شود. روال کار به این صورت است که، نخست پرسش و سند را به صورت مستقل با دو کدگذار برت کدگذاری می‌کنند و امتیاز ارتباط بین آنها، به صورت زیر محاسبه می‌شود.

پرسش: q بردار متناظر با پرسش: (‏طول n)‏

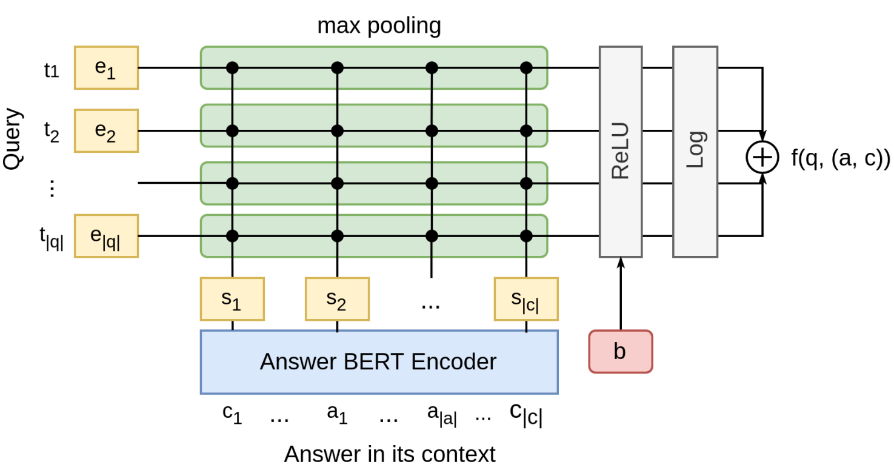
سند: d بردار متناظر با سند: (‏طول m)

|  |  |
| --- | --- |
| رابطه 2-19 |  |

سپس، ColberT ابتدا امتیاز هر نشانه جاسازی شده پرسش را بر روی تمامی اسناد محاسبه می‌کند، و همه این امتیازات را به عنوان امتیاز ارتباط نهایی بین q و d جمع می‌کند. این روش برای معیار EM در سه مجموعه داده NQ، TQ و SQ به ترتیب مقادیر 6/48، 2/70 و 8/51 را کسب کرده است.

به عنوان مثال دیگر از به کارگیری بازیاب مبتنی بر تعامل-بازنمایی، می­توان به مدل معرفی شده توسط ژائو[[153]](#footnote-153) و همکاران در سال 2020 [82] تحت عنوان SPARTA اشاره نمود. آنها در این روش، با استفاده از ضرب داخلی بین یک پرسش کدگذاری شده بدون زمینه[[154]](#footnote-154) (‏برای مثال، جاسازی کلمات با برت) ‏و یک سند کدگذاری شده مبتنی بر زمینه[[155]](#footnote-155) (‏برای مثال، کدگذار برت)، ‏ یک طبقه­بندی کننده عصبی را برای محاسبه میزان تطابق نشانه­ها در پرسش و اسناد توسعه می­دهد. با توجه به بازنمایی­های پرسش و سند، وزن هر نشانه پرسش با max-pool، ReLU و log محاسبه می‌شود؛ امتیاز تطابق نهایی مجموع اوزان نشانه­های پرسش است. معماری این روش در شکل 2-19 قابل مشاهده می­باشد. روش SPARTA توانسته برای معیارهای F1 و EM در مجموعه داده­ OpenSQuAD به ترتیب مقادیر 5/66 و 3/59 و در مجموعه داده OpenNQ مقادیر 8/36 و 5/37 را کسب نماید.

روش بازنمایی-تعامل به دلیل تبادل خوب بین اثربخشی و کارایی، یک روش امیدوارکننده برای بازیابی متراکم است. اگرچه روش مذکور بسیار مؤثر است، اما بازیابی، زمانی که به اسناد در مقیاس بزرگ اعمال ‌شود، از بار محاسباتی سنگینی رنج می‌برد. به منظور سرعت بخشیدن به محاسبات، برخی پیشنهادها برای محاسبه و ذخیره‌سازی بازنمایی­های تمامی اسناد به صورت آفلاین [۱12]‏، [113]‏، [114]‏، [11۵]‏، [49]‏ ارائه شده است. به این ترتیب، این بازنمایی‌ها زمانی که محاسبه شوند تغییر نخواهند کرد، که به این معنی است که اسناد به طور مستقل از پرسش، کدگذاری می‌شوند و بدین ترتیب تاثیر بازیابی کم­رنگ­تر خواهد بود.



شکل 2-19- معماری روش SPARTA [82]

### 2-6-1-1-3- بازیاب تکرارگر

در بازیابی تکرارگر جستجوی اسناد مرتبط با یک پرسش از یک مجموعه بزرگ در چندین گام انجام می­شود که به آن بازیابی چند مرحله‌ای[[156]](#footnote-156) نیز گفته می‌شود. این موضوع در چند سال گذشته به طور گسترده مورد بررسی قرار گرفته ‌است [113]، [11۵]، [140]، [141]، [142]، [143]، [144]، [145]‏. این بازیابی، به خصوص برای پاسخ به پرسش­های پیچیده‌ای که نیازمند استدلال چند امیدی[[157]](#footnote-157) هستند [146] و [147]، بسیار پر کاربرد است‏. به منظور به دست آوردن تعداد کافی از اسناد مرتبط، پرس­وجوهای جستجو در گام­های مختلف تغییر می­کنند و براساس اطلاعات موجود در گام قبل مجددا فرمول‌بندی می­شوند. مراحل بازیاب تکرارگر عبارتند از: الف) ‏بازیابی اسناد: مشتمل بر تکنیک‌های بازیابی اطلاعات مورد استفاده در هر مرحله بازیابی؛ ب) ‏فرمول­بندی مجدد پرس­وجو: مشتمل بر مکانیزم مورد استفاده برای تولید پرس­وجو برای هر بازیابی؛ ج) ‏مکانیزم توقف بازیابی: روشی برای تصمیم‌گیری در مورد زمان اتمام فرآیند بازیابی. در ادامه هریک از این مراحل شرح داده می­شود.

**الف) بازیابی اسناد:** نخست، تکنیک‌های بازیابی اطلاعات مورد استفاده برای بازیابی اسناد در هر گام بازیابی، با در نظر گرفتن یک پرس و جو مشخص، مرور می­شود. برخی از تحقیقات [140]‏، [142]‏، [145]‏ به صورت تکرار شونده از بازیابی تنک استفاده می‌کنند، و برخی از تحقیقات [113]‏، [11۵]‏، [141]‏، [144]‏ نیز به صورت تکرار شونده از بازیابی متراکم استفاده می‌کنند. در میان کارهایی که با استفاده از بازیابی تنک انجام می‌شود، می­توان به سیستم پرسش و پاسخ پیشنهاد شده توسط کیوئی[[158]](#footnote-158) و همکاران 2019 [‏140] اشاره نمود که در کار پیشنهادی خود، از بازیاب GOLDEN استفاده می­کنند که در این بازیاب، از بازیاب BM25 استفاده می­شود. همچنین می­توان به سیستم پیشنهادی مین[[159]](#footnote-159) و همکاران در سال 2019 [145] و سیستم پیشنهادی لی و همکاران در سال 2006 [86] اشاره نمود که با استفاده از TF-IDF، k سند برتر را بازیابی می‌کنند.

بازیاب مسیر[[160]](#footnote-160) یک بازیاب ساخته شده از شبکه عصبی بازگشتی (‏RNN)‏ است که توسط اسایی[[161]](#footnote-161) و همکاران در سال 2020 [‏144]‏ پیشنهاد شد. این بازیاب، برای یادگیری، مسیرهای استدلال را برای یک پرسش در یک گراف ویکی‌پدیا (گرافی که برای مدل­سازی روابط بین پاراگراف‌های مبتنی بر ابرلینک‌ها و ساختارهای مقالات ویکی‌پدیا ساخته شده‌است) توسعه می‌دهد.

**ب) فرمول­سازی مجدد پرس­وجو:** به منظور به دست آوردن مقدار کافی از اسناد مرتبط، پرس­وجوهای متنوعی براساس پرس­وجوی قبلی و اسناد بازیابی شده برای هر مرحله از بازیابی، تولید می‌شوند. پرس­وجوهای تولید شده دو حالت دارند. 1) ‏فرم صریح، برای مثال پرس­وجوهای زبان طبیعی [140]، [142]، [143]‏؛ و ۲)‏ فرم ضمنی، یعنی نمایش متراکم [113]‏، [11۵]‏، [141]‏.

برخی کارها یک پرس­وجوی جدید به شکل زبان طبیعی تولید می‌کنند. به عنوان مثال، بازیاب GOLDEN ‏وظیفه فرمول­بندی مجدد پرس­وجو را به عنوان یک کار ماشین درک مطلب در نظر می­گیرد، زیرا هر دوی آنها یک پرسش و تعدادی سند را به عنوان ورودی در نظر می‌گیرند و هدف آنها تولید یک رشته در زبان طبیعی است. هدف از فرموله کردن مجدد پرس­وجو در ماشین درک مطلب، ایجاد یک پرس­وجوی جدید، به جای تلاش بر یافتن پاسخ است، که به یافتن اسناد با پشتیبانی بیشتر در مرحله بازیابی بعدی کمک می‌کند.

مائو و همکاران در سال 2020 [143]، یک ماژول گسترش پرس­وجو (GAR) را با استفاده از مدل پایانه به پایانه از پیش آموزش‌دیده BART [148] توسعه می‌دهند، که پرسش اولیه را به عنوان ورودی می‌گیرد و پرس­وجوهای جدید تولید می‌کند. این ماژول با استفاده از خروجی­های حاوی پاسخ در سطوح مختلف، جمله‌ای که پاسخ به آن تعلق دارد، و عنوان قطعه‌ای که حاوی پاسخ است، آموزش داده می‌شود. در این روش، معیار کارایی EM در مجموعه داده­های NQ و TriviaQA به ترتیب مقادیر 3/45 و 2/62 را کسب کرده است.

برخی کارهای دیگر، بازنمایی‌های متراکمی را تولید می‌کنند که برای جستجو در یک فضای پنهان[[162]](#footnote-162) مورد استفاده قرار می‌گیرند. به عنوان مثال، سیستم ارائه شده توسط داس[[163]](#footnote-163) و همکاران در سال 2019 [‏113]‏، که یک استدلال کننده چند مرحله­ای[[164]](#footnote-164) پرسش ورودی و بازنمایی‌های مخفی در سطح نشانه (توکن) را که با استفاده از ‏GRU[[165]](#footnote-165) ‏[149]‏ از خواننده بدست آمده، به عنوان ورودی می­گیرد و با استفاده از یادگیری تقویتی، یک بردار پرس­وجوی جدید را آموزش می­دهد. این روش در معیار کارایی F1 بر روی مجموعه داده­های QUASAR-T، SearchQA و TriviaQA به ترتیب به مقادیر 3/46، 1/61 و 03/61 دست یافته است.

در مقایسه، پرس­وجوی صریح به راحتی برای انسان قابل ‌درک و قابل‌کنترل است اما توسط اصطلاحات موجود در واژگان محدود می‌شود، در حالی که پرس و جوی ضمنی در یک فضای معنایی تولید می‌شود که می‌تواند از محدودیت واژگان خلاص شود اما قابلیت تفسیر ندارد.

**ج) مکانیزم توقف بازیابی:** روش بازیابی تکرارگر، امکان بیشتری برای جمع‌آوری متون مناسب‌تر در اختیار می­گذارد، اما کارایی بازیابی با افزایش تعداد تکرارها، به طور چشمگیری کاهش می‌یابد. با توجه به مکانیزم توقف بازیابی تکرارگر، اکثر سیستم‌های موجود علاقه­مند به تعیین تعداد ثابتی از تکرارها [‏113]‏، [140]‏، [141]‏، [142]‏، [145]‏ و یا حداکثر تعداد اسناد بازیابی شده [11۵]‏، [143]‏ می­باشند، که باعث تضمین کارایی می­شود. در سیستم پرسش و پاسخ پیشنهاد شده توسط کارتزوالد[[166]](#footnote-166) و همکاران در سال 2018 [131]، استدلال می­شود که وجود یک تعداد ثابت از اسناد برای همه پرسش­های ورودی بهینه نیست و به جای آن، یک بازیابی تطبیقی اسناد در DrQA [41]‏ را پیشنهاد می­دهد. آنها دو روش پویا برای تعیین تعداد اسناد بازیابی شده برای هر پرس­و جو پیشنهاد می‌کنند، 1) روش ساده ابتکاری مبتنی بر آستانه[[167]](#footnote-167) و 2) یک طبقه‌بندی کننده قابل آموزش با استفاده از رگرسیون ریدج ترتیبی[[168]](#footnote-168). از آنجایی که مشخص کردن تعداد تکرارها برای پرسش­هایی که نیازمند استدلال هستند، دشوار است، بازیاب مسیر[[169]](#footnote-169) [144]‏ تنها زمانی به بازیابی آن پایان می‌دهد که نشانه پایان شواهد[[170]](#footnote-170) (‏برای مثال [EOE]‏) ‏توسط بازیاب بازگشتی[[171]](#footnote-171) شناسایی شود.

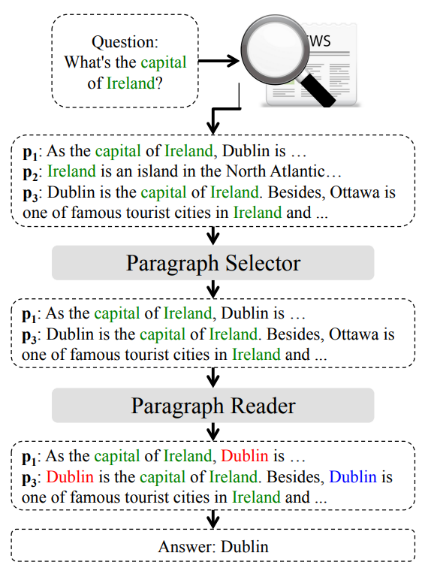
علاوه بر این، معمولاً سیستم‌های بازیابی اطلاعات دو فاکتور بهینه‌سازی یعنی دقت[[172]](#footnote-172) و فراخوانی[[173]](#footnote-173) را در نظر می­گیرند. اولی نسبت اسناد مرتبط بازگشتی به تعداد کل اسناد بازگشتی را محاسبه می‌کند در حالی که دومی تعداد اسناد مرتبط بازگشتی به تعداد کل اسناد مرتبط در مخزن[[174]](#footnote-174) را در نظر می­گیرد. با این حال، برای سیستم‌های پرسش و پاسخ، فاکتور فراخوانی بسیار مهم‌تر از دقت است زیرا پس پردازش معمولا بر اسناد بازگشتی اعمال می‌شود [‏150]‏.

### 2-6-1-2- پس پردازش اسناد

پس­پردازش اسناد، بخش بسیار مهمی در فرآیند پاسخگویی است، زیرا اسناد بازیابی شده به ناچار شامل اسناد غیر مرتبط هستند، و گاهی اوقات، تعداد اسناد بازیابی شده به شدت زیاد است که می­تواند خواننده را تحت تاثیر قرار دهد. همان­طور که در بخش­­های پیشین معرفی شد، پس­پردازش اسناد در معماری سیستم­های پرسش و پاسخ مدرن مشابه با روش سنتی است. هدف از این روش کاهش تعداد اسناد کاندید و تنها صدور مجوز ورود به مرتبط‌ترین اسناد به مرحله بعدی است.

در چند سال گذشته، این ماژول به صورت گسترده مورد بررسی قرار گرفته‌است [‏42]‏، [47]‏، [48]‏، [133]‏، [151]‏، [152]‏، [153]‏، [154]‏. به عنوان مثال، سیستم ارائه شده توسط نیشیدا[[175]](#footnote-175) و همکاران [‏۲۸]‏ که در بخش 2-6-1-1-2 مورد بررسی قرار گرفت که در آن یک رتبه‌بندی کننده اسناد عصبی به همراه خواننده را از طریق یادگیری تقویتی (‏RL) ‏آموزش می‌دهند.

همچنین لین[[176]](#footnote-176) و همکاران در سال 2018 [47]، سیستم DS-QA را پیشنهاد کردند که در آن‏ با اندازه‌گیری احتمال حاوی پاسخ بودن هر پاراگراف، پاراگراف­های نویز از اسناد بازیابی شده حذف می­شدند. همانگونه که در شکل 2-20 نیز قابل مشاهده است، این روش از دو بخش اصلی تشکیل شده است که عبارتند از:



2-20- شمای کلی روش DS-QA [47]

الف) انتخاب­گر پاراگراف، که پاراگراف­های محتمل برای پاسخ را انتخاب می­کند و ب) خواننده پاراگراف، که پاسخ را از پاراگراف­های منتخب، استخراج می­کند. این روش در معیارهای کارایی F1 و EM در مجموعه داده­های Quasar-T و SearchQA مورد آزمایش قرار گرفته اند. در مجموعه داده Quasar-T معیار F1 و EM به ترتیب مقادیر 3/71 و 7/60، و در مجموعه داده SearchQA معیارهای مذکور مقادیر 2/81 و 1/85 را به ترتیب، کسب کرده­اند.

اسکوتن[[177]](#footnote-177) و همکاران در سال 2009 [‏۹۶] ‏به بررسی دو گروه رتبه­بندی کننده اسناد می‌پردازند که بر اساس احتمال پاسخگویی به پرسش داده‌ شده، امتیازهایی را به اسناد بازیابی شده اختصاص می‌دهد. که عبارتند از: الف) رتبه­بندی­کننده InferSent ، که یک شبکه عصبی پیشرو است که از بازنمایی­های جملات بدست آمده از روش InferSent ‏[۱55] و شباهت معنایی،‏ برای رتبه‌بندی اسناد استفاده می‌کند. ب) رتبه‌بندی کننده شبکه - ارتباط[[178]](#footnote-178) [۱56]‏، که بر اندازه‌گیری ارتباط بین کلمات در پرسش و اسناد تمرکز می‌کند. آزمایش‌ها نشان می‌دهد که فاکتور ارتباط کلمات به طور قابل‌توجهی عملکرد بازیابی را بهبود می‌بخشد و فاکتور شباهت معنایی برای عملکرد کلی سودمندتر است.

یون[[179]](#footnote-179) و همکاران در سال 2018 [48]‏، یک رتبه­بندی کننده پاراگراف را با استفاده از دو RNN جداگانه که از معماری کدگذار دوگانه پیروی می­کنند، توسعه می‌دهند. هر جفت پرسش-سند، به عنوان ورودی رتبه­بندی کننده داده می‌شود تا بازنمایی‌های خود را به صورت مستقل به دست آورد و از ضرب داخلی برای محاسبه امتیاز ارتباط استفاده می‌شود. در این روش معیار EM در مجموعه داده SQuAD مقدار 2/30 را کسب نموده است.

### 2-6-1-2- خواننده

خواننده، یک مؤلفه اصلی دیگر یک سیستم پرسش و پاسخ مدرن و همچنین ویژگی اصلی است که سیستم‌های پرسش و پاسخ را از دیگر سیستم‌های بازیاب اطلاعات یا استخراج­گر اطلاعات متمایز می‌سازد. خواننده معمولاً به صورت یک مدل درک مطلب ماشین عصبی پیاده­سازی می‌شود. هدف آن استنتاج پاسخ از مجموعه‌ای از اسناد منظم است، و در مقایسه با وظیفه اصلی یک مدل درک مطلب ماشین عصبی که در آن در اکثر موارد، تنها یک قطعه از متن داده می‌شود، چالش برانگیزتر است [‏۱19]، [157]، [158]‏. به طور کلی، خوانندگان موجود را می­توان به دو نوع طبقه‌بندی کرد: خواننده استخراج­گر که محدوده پاسخ را از اسناد بازیابی شده پیش‌بینی می‌کند، و خواننده مولّد که پاسخ‌ها را با استفاده از مدل‌های توالی به توالی[[180]](#footnote-180) ‏در زبان طبیعی تولید می‌کند. اکثر سیستم‌های پرسش و پاسخ پیشین از یک خواننده استخراج­گر [‏41]‏، [۱12]‏، [113]‏، [44]‏، [47]‏، [11۵]‏، [132]‏، [145]‏ استفاده می­کنند، در حالی که برخی از سیستم‌های اخیر از خواننده­های مولّد [50]‏، [141]‏، [۱59]‏ استفاده می­کنند.

### 2-6-1-2-1- خواننده استخراج­گر

خواننده استخراج­گر براساس این فرض استوار است که پاسخ صحیح به پرسش داده‌ شده قطعاً در اسناد وجود دارد، و معمولاً بر یادگیری برای پیش‌بینی موقعیت شروع و پایان یک پاسخ از اسناد بازیابی شده تمرکز می‌کند. روش­های موجود در این حوزه با توجه به اینکه آیا اسناد بازیابی شده به طور مستقل پردازش می‌شوند یا به طور مشترک، تقسیم­بندی می­شوند.

بسیاری از سیستم‌های قبلی [‏29]‏، [47]‏، [142]‏، [144]‏، [145]‏ اسناد بازیابی شده را بر اساس احتمال وجود پاسخ رتبه‌بندی می‌کنند و محدوده پاسخ را از الحاق پرسش و محتمل‌ترین سند(‏ها) ‏استخراج می‌کنند. برای مثال، سیستم پرسش و پاسخDS -QA [47]‏ که در بخش 2-6-1-2 بررسی شد، یک محدوده پاسخ را از پاراگراف انتخاب‌شده توسط ماژول انتخاب­گر پاراگراف، استخراج می‌کند. همچنین سیستم DPR [29] که در بخش 2-6-1-1-2 بررسی شد، ‏احتمال حاوی پاسخ بودن یک متن و یک نشانه موقعیت آغاز و پایان محدوده پاسخ را با استفاده از خواننده مبتنی بر برت محاسبه، و با ترکیب آنها، پاسخ با بالاترین احتمال را انتخاب می‌کند.

برخی سیستم‌ها از یک خواننده مبتنی بر گراف [144]‏، [145]‏، برای یادگیری استخراج محدوده پاسخ استفاده می­کنند. که در ادامه به آنها پرداحته شده است.

مین و همکاران در سال 2019 [145]، سیستمی پیشنهاد کردند که با استفاده از یک خواننده گراف پاسخ­ها را استخراج می­نماید. روال کار به این صورت است که سیستم در ابتدا یک گراف را به عنوان ورودی گرفته و با استفاده از شبکه‌های کانولوشن گراف [‏۱60]‏ بازنمایی­های متن را یاد می‌گیرد و سپس پاسخ را از محتمل‌ترین محدوده استخراج می‌کند. این روش بر روی مجموعه داده TriviaQA به مقدار 56 درصد برای معیار F1 دست یافته است.

اسائی و همکاران در سال 2020 [144]، سیستمی پیشنهاد کردند که با استفاده از بازیاب مسیر[[181]](#footnote-181) وبکارگیری خواننده برت، طور همزمان مسیرهای استدلال را رتبه‌بندی می­کند و محدوده پاسخ را از محتمل­ترین کاندیدهای پاسخ، که با استفاده از یادگیری چند وظیفه‌ای بدست آمده، استخراج می­نماید. با این وجود، مدل نمی­تواند از نشانه­های مختلف موجود در اسناد برای استخراج پاسخ بهره ببرد، زیرا اسناد بازیابی شده به طور مستقل پردازش می شوند، و در نتیجه منجر به تضعیف کارایی در عملکرد به ویژه در مواردی که پرسش‌های داده‌شده نیاز به استدلال چند امیدی دارند، می­شود. این روش توانسته بر روی مجموعه داده SQuAD به مقدار 4/72 برای معیار F1 و 5/59 برای معیار EM دست یابد.

در مقابل، برخی از سیستم­ها [41] ، [132] ، [82] ، [161]، مشترکاً بر اساس کل اسناد بازیابی شده، یک بازه پاسخ را استخراج می­کنند. برای مثال، سیستم DrQA [‏41] که در بخش 2-5 بررسی شد،‏ اسناد بازیابی شده را به پاراگراف‌ها تجزیه می‌کند و ویژگی‌های مختلفی ازجمله، بخش گفتار (‏POS)‏، موجودیت نام گذاری شده (‏NE)‏، فرکانس اصطلاحات (‏TF)‏ و غیره را استخراج می‌کند. سپس خواننده DrQA، که با یک Bi-LSTM چندلایه‌ای پیاده­سازی شده است، پرسش و پاراگراف‌ها را به عنوان ورودی گرفته و محدوده پاسخ را پیش‌بینی می‌کند. در این فرآیند، به منظور ایجاد قابلیت مقایسه در پاسخ­ها، از تابع نمایی نرمال نشده استفاده می‌شود و argmax را برای تمام بازه­های پاسخ محاسبه می­کند تا نتیجه نهایی را بدست آورد.

سیستم BERTserini نیز با حذف لایه Softmax، خواننده خود را براساس برت توسعه می‌دهد تا بتواند امتیازهای پاسخ­ها را در میان پاراگراف‌های مختلف مقایسه و جمع کند.

کلارک[[182]](#footnote-182) و همکاران در سال 2018 [‏161]‏، ادّعا کردند که استفاده از امتیازات نرمال نشده (‏به عنوان مثال، امتیازات نمایی یا لگاریتمی)‏ برای محدوده‌های پاسخ، بهینه نیست و یک مکانیزم مشترک نرمال­سازی را به منظور اصلاح تابع هدف برای نرمال­سازی امتیازات شروع و پایان در تمام پاراگراف‌ها، پیشنهاد کردند. سیستم پیشنهادی آنها در مجموعه داده SQuAD، مقدار 37/72 را برای معیار F1 و 08/64 را برای معیار EM کسب کرده است.

همچنین می­توان بسیاری از سیستم‌های پرسش و پاسخ دیگر مانند [30]‏، [113]‏، [115]‏، [140]‏، [133]‏، [82]‏ را نام برد که خوانندگان خود را با استفاده از این مکانیزم و بر اساس مدل‌های درک مطلب ماشین عصبی توسعه می‌دهند.

### 2-6-1-2-2- خواننده مولّد

خواننده مولد به جای استخراج محدوده پاسخ که معمولاً بر مدل‌های دنباله به دنباله متکی هستند، تا حد امکان پاسخ‌های طبیعی، تولید می­کند. برای مثال، تان[[183]](#footnote-183) و همکاران در سال 2018 [162]، مدلی تحت عنوانS-Net ‏ را پیشنهاد کردند که در آن‏ روش‌های استخراج و تولید را ترکیب می­نمایند تا یکدیگر را کامل کنند. این روش از یک مدل استخراج شواهد[[184]](#footnote-184) برای پیش‌بینی مرزهای مهم پاسخ در متون استفاده می‌کند و سپس آنها را به عنوان ورودی به یک مدل سنتز پاسخ پایانه به پایانه می­دهد تا پاسخ نهایی را تولید کند. معماری کلی این روش در شکل 2-21 قابل مشاهده می­باشد.

Synthesis & Generation

Extraction

Question

Answer

Evidance

Snippet

Evidance

Extraction

Passage

شکل 2-21- معماری مدل S-Net [107]

این مدل از تابع زیان log-likelihood منفی برای بهبینه­سازی بهره می­برد و در معیارهای کارایی ROUGE-L و BLEU-1 در مجموعه داده MS-MARCO به ترتیب به مقادیر 65/46 و 78/44 دست یافته است.

اخیراً، برخی از سیستم‌های پرسش و پاسخ،‏ [50]‏، [141]‏، [۱59]‏ از مدل‌های زبانی پایانه به پایانه از پیش آموزش‌دیده برای توسعه خوانندگان خود، مانندBART ‏ [148]‏ و T5 [‏163]‏، استفاده می‌کنند. به عنوان مثال، لوئیس[[185]](#footnote-185) و همکاران در سال 2020 [50]، سیستم RAG را پیشنهاد کردند که از یک خواننده مولّد استفاده می­کرد. در واقع این سیستم‏ یک مدل BART از پیش آموزش‌دیده را به عنوان خواننده‌اش در نظر گرفته و برای تولید پاسخ، پرسش ورودی و همچنین اسناد بازیابی شده توسط DPR [29] را به عنوان ورودی می­گیرد. در این روش معیارهای کارایی EM در مجموعه داده­های NQ، TQA و WQ به ترتیب دارای مقادیر 9/44، 8/55 و 44 می­باشد. همچنین ایزاکارد[[186]](#footnote-186) و همکاران در سال 2020 [159]، سیستم ‏FID[[187]](#footnote-187) را پیشنهاد کردند که از دو بخش اصلی تشکیل شده است. بخش نخست (بخش بازیابی)، که در آن از دو روش مبتنی بر BM25 و DPR استفاده می­شود. و بخش دوم (بخش خواننده مولّد)، که ابتدا هر سند بازیابی شده را به طور مستقل با استفاده از کدگذار T5 یا BART کدگذاری کرده و سپس برای تولید پاسخ نهایی، مکانیزم توجه را با استفاده از کدگشا، بر روی تمام بازنمایی‌های خروجی به کار می­گیرد. با این وجود، نتایج بدست آمده اغلب از خطای نحوی، عدم انسجام یا منطق [۱64]‏ رنج می‌برند. در این روش، معیار EM بر روی دو مجموعه داده Natural Question و TriviaQA در بهترین حالت، به ترتیب به مقادیر 5/46 و 7/64 دست یافته است.

### 2-6-1-3- پس پردازش پاسخ

در سال‌های اخیر، تکنیک‌های درک مطلب ماشین عصبی به سرعت در حال پیشرفت بوده‌اند، اما اکثر مدل‌های موجود، قادر به استخراج پاسخ، تنها از یک یا چند متن کوتاه می­باشند و در مواردی که پاسخ صحیح از شواهد مختلف در یک سند بزرگ یا چند سند می‌آید، شکست می‌خورند [‏۱65]‏. ماژول پس پردازش پاسخ، برای کمک به تشخیص پاسخ نهایی از مجموعه‌ای از نامزدهای پاسخ استخراج‌شده توسط خواننده، توسعه داده شده ‌است. روش‌های اتخاذ شده در سیستم‌های موجود را می توان به دو دسته تقسیم کرد: روش مبتنی بر قانون [48]‏، [۱65]‏ و روش مبتنی بر یادگیری [‏130]‏، [۱65]. به عنوان مثال، ونگ[[188]](#footnote-188) و همکاران در سال 2018 [۱65]‏ برای تصمیم‌گیری در مورد پاسخ نهایی، دو رتبه­بندی کننده پاسخ، (یک "رتبه­بندی کننده مبتنی بر توان[[189]](#footnote-189)" و یک "رتبه­بندی کننده مبتنی بر پوشش[[190]](#footnote-190)") را برای جمع‌آوری شواهد از متون مختلف پیشنهاد کردند. " رتبه­بندی کننده مبتنی بر توان" یک روش مبتنی بر قانون است که محاسبات احتمالی را براساس کاندیدهای پیش‌بینی‌ شده، انجام می‌دهد و به هیچ آموزشی نیاز ندارد. "رتبه­بندی کننده مبتنی بر پوشش" با استفاده از مدل تطابق مبتنی بر توجه LSTM توسعه داده شده ‌است [166]‏. در این رتبه­بند، ابتدا بخش‌هایی را که حاوی پاسخ­های مشابه هستند را در یک شبه متن[[191]](#footnote-191) به هم متصل می‌کند و سپس از شبه متن به دست آمده پاسخ را استخراج می­کند. آزمایش‌ها نشان می‌دهند که یک ترکیب وزن­دار از خروجی‌های رتبه­بندی کننده­های مختلف، بهترین عملکرد را در مجموعه داده­های Quasar-T، SearchQA و TriviaQA دارد. در مجموعه داده­های مذکور، معیار EM به ترتیب مقادیر 3/42، 0/57 و 6/50 و معیار F1 به ترتیب مقادیر 6/49، 2/63 و 3/57 را کسب کرده­اند. کارتزوالد و همکاران در سال 2019 [130]، سیستم RankQA را که ‏یک ماژول رتبه‌بندی پاسخ است، توسعه می‌دهند. این سیستم، شامل سه مرحله است: استخراج ویژگی، تجمیع پاسخ و رتبه‌بندی مجدد. در مرحله نخست، K کاندید پاسخ برتر بدست آمده از خواننده را به عنوان ورودی در نظر گرفته، سپس مجموعه‌ای از ویژگی‌ها را از بازیاب و خواننده استخراج می‌شود، از بازیاب ویژگی­هایی مانند شباهت سند-پرسش، طول پرسش و طول پاراگراف، و از خواننده ویژگی­هایی مانند امتیاز واقعی کاندیدهای پاسخ، برچسب‌های گفتار پاسخ و موجودیت نام­گذاری شده پاسخ. در مرحله دوم، همه کاندیدها در یک محدوده پاسخ یکسان، برای تولید مجموعه‌ای از ویژگی‌های تجمعی مانند جمع، میانگین، حداقل و حداکثر امتیازات محدوده و غیره، گروه­بندی می­شوند. در نهایت در مرحله سوم، با استفاده از ویژگی‌های مذکور، از یک شبکه رتبه‌بندی مجدد، مانند یک شبکه پیش‌خور[[192]](#footnote-192) یا یک RNN، برای یادگیری رتبه‌بندی مجدد پاسخ‌ها (جهت انتخاب بهترین پاسخ به عنوان پاسخ نهایی) استفاده می­شود. در این روش معیار کارایی EM برای چهار مجموعه داده SQuADOPEN، CuratedTREC، WebQuestions و WikiMovies به ترتیب مقادیر 2/54، 9/65، 8/53 و 65 دست یافته است.

### 2-6-2- روش‌های پایانه به پایانه

در سال‌های اخیر، سیستم‌های مختلف پرسش و پاسخ [28]‏، [44]‏، [‏49]‏ توسعه‌یافته­اند که در آن‌ها بخش­های بازیاب و خواننده را می‌توان به صورت پایانه به پایانه آموزش داد. علاوه بر این، در برخی سیستم‌ها بخش خواننده وجود ندارد و تنها بخش بازیاب [136]‏ وجود دارد، و همچنین برخی دیگر سیستم­های پرسش و پاسخ وجود دارند که قادر به پاسخگویی به پرسش­ها، بدون مرحله بازیابی هستند، که اغلب، مدل‌های زبان توالی به توالی از پیش آموزش‌دیده هستند [148]‏، [۱63]‏، [۱67]‏، [۱68]‏. در ادامه، این سه نوع سیستم، یعنی بازیاب-خواننده[[193]](#footnote-193)، بازیاب-تنها[[194]](#footnote-194) و بازیاب-آزاد[[195]](#footnote-195) را معرفی خواهند شد.

### 2-6-2-1- سیستم­های بازیاب – خواننده

تکنیک‌های یادگیری عمیق، بخش­های بازیاب و خواننده را در یک سیستم پرسش و پاسخ قادر می‌سازد تا به صورت پیوسته قابل آموزش باشند [‏28]، [44]، [50]، [49]‏. به عنوان مثال، در مورد پیشنهاد نیشیدا و همکاران [‏28]‏ همانگونه که در بخش 2-6-1-1-2 نیز بررسی شد، پیشنهاد هم­زمان آموزش مشترک بازیاب و خواننده را با استفاده از یادگیری چند وظیفه‌ای براساس مدل BiDAF ‏ [37]‏ را می­دهد، سپس تشابه یک متن به پرسش داده ‌شده محاسبه شده و در نهایت موقعیت آغاز و پایان محدوده پاسخ پیش‌بینی می‌شود. معیار کارایی EM در مجموعه داده CNN مقدار 9/76 و در مجموعه داده DailyMail 6/79 گزارش شده است. همچنین گو و همکاران نیز در پیشنهاد مدل REALM ‏ [44]‏ خود (بر اساس توضیحات ارائه شده در بخش 2-5)، یک سیستم پرسش و پاسخ از نوع بازیاب-خواننده ارائه کردند. این مدل، یک مدل زبان پوششی از پیش آموزش‌دیده است که شامل یک بازیاب و یک خواننده عصبی است که قادر به محاسبه گرادیان، با توجه به پارامترهای مدل و پس انتشار گرادیان در سراسر شبکه است. از آنجا که هر دو ماژول با استفاده از شبکه‌های عصبی توسعه داده شده‌اند، سرعت پاسخ به یک پرسش مهم‌ترین مساله در طول استنتاج، به خصوص در مجموعه بزرگ از اسناد است.

### 2-6-2-2- سیستم­های بازیاب – تنها

برای افزایش کارایی سیستم پرسش و پاسخ، برخی از سیستم‌ها تنها با اتخاذ یک بازیاب توسعه می‌یابند و بخش خواننده را که معمولاً زمان­برترین مرحله است، حذف می‌کنند. سئو و همکاران در سال 2019 [136]، سیستم DenSPI را پیشنهاد کردند که با توجه به مجموعه‌ای از اسناد مانند مقالات ویکی‌پدیا، شاخص­گذاری عبارات را به صورت آفلاین ایجاد می‌کرد. در این شاخص، هر عبارت کاندید از پیکره، با الحاق دو بردار، یعنی یک بردار تنک (‏برای مثال tf - idf) ‏و یک بردار متراکم (‏برای مثال، کدگذار برت) ‏نشان داده می‌شود. برای استنتاج، پرسش داده‌ شده به همان روش کدگذاری می‌شود و FAISS [۱68]‏ برای جستجوی عبارات مشابه به عنوان پاسخ نهایی به کار گرفته می‌شود. آزمایش‌ها نشان می‌دهند که این روش به دستاوردهای بهره‌وری قابل‌ توجهی دست می‌یابد و در حالی که دقت را حفظ می‌کند، هزینه‌های محاسباتی را به طور قابل ‌توجهی کاهش می‌دهد. با این وجود، از آنجایی که سیستم، تشابه بین هر عبارت و پرسش را به طور مستقل محاسبه می‌کند، منجر به نادیده گرفته شدن اطلاعات متنی که معمولاً برای پاسخ به پرسش­ها، حیاتی هستند، می­شود. نتایج این روش در مجموعه داده SQuAD مقادیر 3/42 برای معیار F1 و 4/33 برای معیار EM می­باشد.

### 2-6-2-3- سیستم­های بازیاب – آزاد

پیشرفت اخیر در مدل‌های زبان پیش آموزش شده توالی به توالی مانند GPT-2 [۱67]‏، GPT-3 [۱68]‏، BART [148]‏ و T5 [۱63]‏ موجی از بهبودهایی را برای NLG[[196]](#footnote-196)به ارمغان آورده است، که بیشتر آنها با استفاده از معماری‌های مبتنی بر تبدیل[[197]](#footnote-197) ساخته شده‌اند. مطالعات صورت گرفته [‏۱69]، [۱70]‏ نشان می‌دهند که مقدار زیادی از دانش آموخته‌شده از داده‌های متنی در مقیاس بزرگ می‌تواند در پارامترهای خاصی ذخیره شوند، و بنابراین این مدل‌ها قادر به پاسخگویی به پرسش­ها بدون نیاز به دسترسی به دانش خارجی هستند. به عنوان مثال، رَدفورد[[198]](#footnote-198) و همکاران در سال 2019 [167]، مدل GPT-2 را پیشنهاد کردند که قادر است با بکارگیری مدل­های پیش­آموزش شده مانند BART و T5 و بدون بهینه­سازی پارامتری خاصی، پاسخ صحیح را از پرسش زبان طبیعی تولید نماید و عملکرد آن بر روی مجموعه داده SQuAD، برای معیار دقت 1/63 درصد گزارش می­شود. پس از آن، براون[[199]](#footnote-199) و همکاران در سال 2020 [168] مدل GPT-3 را معرّفی کردند که توانست با استفاده از یادگیری چند شات[[200]](#footnote-200) به عملکرد قابل رقابتی در مقایسه با روش‌های تنظیم پارامترهای پیشرفته پیشین دست ‌یابد. این روش در مجموعه داده TriviaQA به دقت قابل قبول 2/71 دست یافته است، و این در حالی است که دقت این روش در مجموعه داده­های WebQS و NaturalQS به ترتیب 9/29 و 5/41 است که بسیار ناامید کننده می­باشد. اخیراً، پترونی[[201]](#footnote-201) و همکاران در سال 2019 [۱70]‏ به طور جامع قابلیت مدل‌های زبانی برای پاسخ به پرسش­ها، بدون دسترسی به دانش خارجی را ارزیابی کرده­اند. آزمایش‌ها نشان می‌دهد که مدل‌های زبانی از پیش آموزش‌دیده، قادر به دستیابی به عملکرد چشمگیر در ­benchmarkهای مختلف هستند و چنین روش‌های بدون بازیابی، رویکرد اساسا متفاوتی را برای ساخت سیستم‌های پرسش و پاسخ ایجاد می‌کنند.

روش­های مختلف اتخاذ شده در سیستم­های پرسش و پاسخ مدرن به صو در جدول 2-2 گردآوری شده است.

## **2-7- علّیت**

مفهوم علّیت، اندازه­گیری و کمّی نمودن آن همواره یکی از چالش­های پیچیده پیش روی پژوهشگران و دانشمندان بوده است. نوربرت وینر[[202]](#footnote-202) در سال1956 [199] نظریه خود را تحت عنوان علّیت وینر مطرح نمود که در آن مفهوم پیشبینی­پذیری را جایگزین مفهوم علّیت کرد و به این ترتیب گام بزرگی در راستای کمّی­سازی مفهوم علّیت برداشته شد. بر اساس نظریه علّیت وینر، در صورتی که افزودن اطلاعات گذشته و حال X به اطلاعات گذشته و حال Y دقت پیش­بینی وضعیت آینده Y را بهبود دهد، X علّت Y در قالب مفهوم علّیت وینر است. با توجه به نظریه علّیت وینر، همواره علّت باید از نظر زمانی مقدم بر معلول باشد و به همین دلیل ارتباط علّی را ارتباط تأخیری نیز می­نامند. علی­رغم اینکه نظریه علّیت وینر، بستر مناسبی برای بررسی ارتباطات علّی است، اما فاقد چارچوبی برای کمّی­سازی ارتباطات علّی موجود در یک شبکه می­باشد. از این­رو چارچوبی سیستماتیک برای کمّی­سازی مفهوم علّیت وینر به کمک مدل­های خودبازگشتی توسط گرنجر[[203]](#footnote-203) ارائه شده است که به علّیت گرنجر [200] معروف است. اساساً دو رویکرد پارامتری و مبتنی بر مدل، دو مورد از پرکاربردترین رویکردهای تشخیص علّیت هستند. در تخمین­گرهای مبتنی بر مدل نمونه­های حاضر سیگنال با استفاده از یک مدل خود بازگشتی چند متغیره خطی از روی نمونه­های گذشته سیگنال پیش­بینی می­شوند و از روی ضرایب این مدل شدت روابط علّی خطی بدست می­آید. نقطه ضعف این روش­ها را می­توان در ثابت بودن مرتبه مدل و همچنین نادیده گرفتن تاثیرات

جدول 2-2- روش­های مختلف اتخاذ شده در سیستم­های پرسش و پاسخ مدرن

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **سیستم** | **دسته** | **بازیاب** | **پس­پردازش اسناد** | **خواننده** | **پس­پردازش پاسخ** |
| DrQA [41] | خط لوله | تنک | - | استخراج­گر | - |
| R3 [42] | خط لوله | تنک | یادگیری تقویتی | استخراج­گر | - |
| DS-QA [47] | خط لوله | تنک | یادگیری نظارت شده | استخراج­گر | - |
| [165] | خط لوله | تنک | - | استخراج­گر | مبتنی بر قانون مبتنی بر آموزش |
| [151] | خط لوله | تنک | یادگیری نظارت شده | استخراج­گر | - |
| رتبه­بند پاراگراف [48] | خط لوله | تنک | یادگیری نظارت شده | استخراج­گر | مبتنی بر قانون |
| RankQA [130] | خط لوله | تنک | - | استخراج­گر | مبتنی بر آموزش |
| BERTserini [132] | خط لوله | تنک | - | استخراج­گر | - |
| Multi-Passage  BERT [133] | خط لوله | تنک | یادگیری انتقالی | استخراج­گر | - |
| [28] | خط لوله | متراکم | - | استخراج­گر | - |
| DPR [29] | خط لوله | متراکم | - | استخراج­گر | - |
| **سیستم** | **دسته** | **بازیاب** | **پس­پردازش اسناد** | **خواننده** | **پس­پردازش پاسخ** |
| ColBERT-QA [30] | خط لوله | متراکم | - | استخراج­گر | - |
| SPARTA [82] | خط لوله | متراکم | - | استخراج­گر | - |
| FID [159] | خط لوله | تنک - متراکم | - | مولد | - |
| Adaptive Retrieval [131] | خط لوله | تکرارگر | - | استخراج­گر | - |
| Multi-step  Reasoner [113] | خط لوله | تکرارگر | - | استخراج­گر | - |
| GOLDEN  Retriever [140] | خط لوله | تکرارگر | - | استخراج­گر | - |
| MUPPET [115] | خط لوله | تکرارگر | - | استخراج­گر | - |
| Path Retriever [144] | خط لوله | تکرارگر | - | استخراج­گر | - |
| Graph Retriever [145] | خط لوله | تکرارگر | - | استخراج­گر | - |
| DDRQA [142] | خط لوله | تکرارگر | - | استخراج­گر | - |
| GAR [143] | خط لوله | تکرارگر | - | استخراج­گر | مبتنی بر قانون |
| **سیستم** | **دسته** | **بازیاب** | **پس­پردازش اسناد** | **خواننده** | **پس­پردازش پاسخ** |
| MDR [141] | خط لوله | تکرارگر | - | استخراج­گر، مولد | - |
| DenSPI [136] | پایانه به پایانه | بازیاب | - | استخراج­گر | - |
| Retrieve -and-Read [46] | پایانه به پایانه | متراکم | - | استخراج­گر | - |
| ORQA [49] | پایانه به پایانه | متراکم | - | استخراج­گر | - |
| REALM [44] | پایانه به پایانه | متراکم | - | استخراج­گر | - |
| RAG [50] | پایانه به پایانه | متراکم | - | مولد | - |

غیرخطی دانست. از سوی دیگر، رویکردهای پارامتری مبتنی بر انتروپی، داده­ها را بصورت مستقیم از روی سیگنال تخمین می­زنند که معمولاً بار محاسباتی سنگینی دارند.

از مفهوم علّیت در سیستم­های پرسش و پاسخ نیز استفاده شده است. برای مثال، ایشیدا[[204]](#footnote-204) و همکاران در سال 2018 [201]، سیستمی برای شناسایی روابط علّی از متون، برای خودکارسازی بازیابی جفت­های پرسش (پرسش­های از نوع چرا) و پاسخ پیشنهاد کردند. در این سیستم، ابتدا جملات حاوی نشانه­های علّی (برای مثال، جملات حاوی کلمات زیرا، چون و ...) از متن استخراج شده و قسمت­های علّت و معلول جمله مشخص می­گردند. به عنوان مثال، جمله (خورشید در شرق طلوع می­کند زیرا زمین به دور محور خود به سمت شرق می­چرخد) را در نظر بگیرید. در این جمله بخش زمین به دور محور خود به سمت شرق می­چرخد، علّت و بخش خورشید در شرق طلوع می­کند، معلول است. لذا باتوجه به اینکه این سیستم پرسش­های از نوع چرا را مد نظر قرار داه است، در جمله مذکور پرسش، جمله چرا خورشید از شرق طلوع می­کند؟ و پاسخ، جمله زیرا زمین به دور محور خود به سمت شرق می­چرخد، می­باشد.

فوکوموتو[[205]](#footnote-205) و همکاران در سال 2007 [202] یک سیستم پرسش و پاسخ را با استفاده از روابط علّی برای پرسش­های از نوع چرا، چگونه، تعریفی و غیرحقیقی[[206]](#footnote-206) پیشنهاد کردند. در روش پیشنهادی آنها روابط معنایی علّی با استفاده از تئوری ساختاری بلاغی[[207]](#footnote-207) مشخص می­شود. سپس بخش­هایی از متون بر اساس روابط علّی بدست آمده بازیابی می­شوند. در نهایت، نامزدهای پاسخ از این قسمت های بازیابی شده حاوی روابط علی استخراج می­شوند. اگر جمله­ای در اسناد با الگوهای استخراج شده مطابقت داشته باشد، به عنوان نامزد پاسخ مناسب در نظر گرفته می­شود.

هیگاشیناکا[[208]](#footnote-208) و ایسوزاکی[[209]](#footnote-209) در سال 2008 [203]، به موضوع اختصاص وزن مناسب به الگوهای دست­ساز پرداختند [204]. آنها نامزدهای پاسخ را بر اساس سه ویژگی که عبارتند از الف) عبارات علّی استخراج شده ازFrameNet [205] و PropBank [206]؛ ب) تشابه محتوا، کلمات مشترک در پرسش و پاسخ های نامزد؛ ج) روابط علّی بدست آمده با استفاده از فرهنگ لغت EDR [207]؛ رتبه­بندی نمودند.

در جدول 2-3 سیستم‌های پرسش و پاسخ مدرن موجود و همچنین رویکردهای اتخاذ شده برای مؤلفه‌های مختلف خلاصه شده است.

## **2-8- چالش‌ها و موانع**

در این بخش، ابتدا چالش‌های کلیدی برای ساخت سیستم‌های پرسش و پاسخ و سپس تحلیل معیارهای موجود که معمولا نه تنها برای سیستم­های پرسش و پاسخ بلکه برای سیستم­های درک مطلب ماشین عصبی نیز لازم است، مورد بحث قرار می­گیرند.

جدول 3-2- خلاصه اطلاعات مقالات مرور شده

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **نام نویسندگان** | **سال انتشار** | **مجموعه داده** | **روش کار** | **ارزیابی** |
| Marius Pasca | 2003 | - | بر اساس معماری سنتی سیستم­های پرسش و پاسخ که شامل سه مرحله آنالیز پرسش، بازیابی سند  و استخراج پاسخ می­باشد. | - |
| S. M. Harabagiu, S. J. Maiorano, and M. A. Paundefinedca | 2003 | - | بر اساس معماری سنتی سیستم­های پرسش و پاسخ که شامل سه مرحله آنالیز پرسش، بازیابی سند  و استخراج پاسخ می­باشد. | - |
| T. Lei, Z. Shi, D. Liu, L. Yang, and F. Zhu | 2018 | MR, SST-1, SUBJ, TREC, MPQA | استفاده از یک مدل مبتنی بر CNN | در معیار دقت مقدار 3/86 درصد را گزارش کرده است |
| W. Xia, W. Zhu, B. Liao, M. Chen, L. Cai, and L. Huang | 2018 | bAbI dataset | استفاده از یک مدل مبتنی بر RNN | در معیار دقت مقدار 9/89 درصد را گزارش کرده است |
| R. Das, S. Dhuliawala, M. Zaheer, and A. McCallum | 2019 | Quasar-T, SearchQA, TRIVIAQA,  SQUAD-open | کدکننده‌های خودشان را آموزش می‌دهند تا هر سند و پرسش را به صورت مستقل در بردارهای متراکم کدگذاری کنند، و امتیاز شباهت بین آن‌ها با استفاده از ضرب داخلی بردارهای آنها (MIPS) محاسبه می‌شود. | برای 4 مجموعه داده ذکر شده و برای 2 معیار F1 و EM به ترتیب مقادیر 97/46، 63/40 برای Quasar-T، 36/61، 26/56 برای SearchQA ، 65/61، 56/61 برای TRIVIAQA و 22/39، 93/31 برای SQUAD-open گزارش شده است. |
| Y. Feldman et al | 2019 | HotpotQA,  SQuAD-Open | کدکننده‌های خودشان را آموزش می‌دهند تا هر سند و پرسش را به صورت مستقل در بردارهای متراکم کدگذاری کنند، و امتیاز شباهت بین آن‌ها با استفاده از ضرب داخلی بردارهای آنها (MIPS) محاسبه می‌شود. | معیار F1 و EM به ترتیب مقادیر 42/40 و 07/31 گزارش شده است. |
| M. J. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi | 2017 | CNN,  DailyMail | اسناد بازیابی شده را از طریق ساختار سلسله مراتبی چند مرحله‌ای، در سطوح (لایه­های) مختلف که شامل یک لایه نگاشت کاراکتر، یک لایه نگاشت کلمه، و یک لایه نگاشت متن است، نمایش می‌دهد، و یک مکانیزم جریان توجه دو طرفه را برای به دست آوردن بازنمایی سند-پرسش به کار می­گیرد. | معیار کارایی EM در مجموعه داده CNN مقدار 9/76 و در مجموعه داده DailyMail 6/79 گزارش شده است |
| A. W. Yu, D. Dohan, M. Luong, R. Zhao, K. Chen, M. Norouzi, and Q. V. Le | 2018 | SQuAD | از CNN و مکانیزم خود-توجهی ‏ و همچنین تعاملات محلی و سراسری برای مدل‌سازی استفاده می‌کند، که به طور قابل ‌توجهی سریع‌تر از مدل‌های تکرار شونده معمول عمل می‌کند. این روش دو نقطه ابتدا و انتهای پاسخ، در یک متن را به عنوان خروجی بر می­گرداند | مقدار 1/51 را برای معیار EM و 6/56 را برای معیار F1 گزارش کرده است. |
| J. Devlin, M. Chang, K. Lee, and K. Toutanova | 2018 | TriviaQA و SQuAD | مرحله بازیابی اسناد، به عنوان متغیر پنهان در نظر گرفته می­شود و در سیستم پرسش و پاسخ ادغام می­گردد. در این حالت، سیستم تنها بر اساس جفت رشته­های پرسش–پاسخ و با استفاده از شبکه برت آموزش می­بیند. | معیار EM در این روش بر روی دو مجموعه داده TriviaQA و SQuAD به ترتیب 1/47 و 2/33 گزارش شده است. |
| D. Chen, A. Fisch, J. Weston, and A. Bordes | 2017 | SQuAD | مدل DrQA، تکنیک بازیابی اطلاعات مبتنی بر TF-IDF و مدل درک مطلب ماشین عصبی را با هم ترکیب کرده و نتایج قابل قبولی در حوزه پاسخ به پرسش­های حقیقی دامنه باز دریافت کرده است. | ارزیابی آن بر روی مجموعه داده SQuAD مقدار 70 درصد برای معیار EM و 79 درصد برای معیار F1 گزارش شده است. |
| K. Guu, K. Lee, Z. Tung, P. Pasupat, and M.-W. Chang | 2020 | NaturalQuestions-Open، WebQuestions و CuratedTrec | مدل REALM، یک مدل زبانی از پیش آموزش‌دیده است که شامل یک بازیاب و یک کدگذار دانش است. در این مدل بازیاب دانش و کدگذار، شبکه‌های عصبی متمایز هستند، که قادر به محاسبه گرادیان با توجه به پارامترهای مدل می­باشند که در سراسر شبکه منتشر می‌شود. | بر روی سه مجموعه داده NaturalQuestions-Open، WebQuestions و CuratedTrec مورد ارزیابی قرار گرفته و معیار EM برای هریک به ترتیب 4/40، 7/40 و 8/46 گزارش شده است. |
| W. Yang, Y. Xie, A. Lin, X. Li, L. Tan, K. Xiong, M. Li, and J. Lin | 2019 | SQuAD | مدل BERTserini، جزئیات متن در سطوح مختلف، از جمله سطح مستند، سطح پاراگراف و سطح جمله به صورت تجربی بررسی می­شوند و نتایج نشان می‌دهد که جزئیات بدست آمده از سطح پاراگراف بهترین کارایی را دارد. | بر روی مجموعه داده SQuAD به مقدار 6/38 برای معیار EM و 1/46 برای معیار F1 دست یافته است. |
| V. Karpukhin, B. O˘guz, S. Min, L. Wu, S. Edunov, D. Chen, and W.-t. Yih | 2020 | NQ ،TriviaQA WQ، TREC ، SQuAD | مدل DPR، از دو رمزگذار برت مستقل استفاده می‌کند، اما مرحله پیش آموزش را به دلیل پر هزینه بودن آن حذف می‌کند و به جای آن، بر یادگیری یک بازیاب قوی با استفاده از جفت­های پرسش و پاسخ‌ متمرکز است | بر روی مجموعه داده­های NQ، TriviaQA، WQ، TREC و SQuAD مورد بررسی قرار گرفته و برای معیار EM به ترتیب مقادیر 5/41، 8/56، 4/42، 4/49 و 1/24 را گزارش کرده است. |
| O. Khattab, C. Potts, and M. Zaharia | 2020 | NQ، TQ و SQ | سیستم پرسش و پاسخ مبتنی ColBERT-QA [30]‏ قسمت بازیاب خود را براساس ColberT [139]‏ توسعه می‌دهد، که در این توسعه به منظور محاسبه امتیاز شباهت، از یک معماری کدگذار دوگانه بر روی بازنمایی­های پرسش و سند استفاده می­شود. | برای معیار EM در سه مجموعه داده NQ، TQ و SQ به ترتیب مقادیر 6/48، 2/70 و 8/51 را کسب کرده است. |
| T. Zhao, X. Lu, and K. Lee | 2020 | OpenSQuAD  OpenNQ، | با استفاده از ضرب داخلی بین یک پرسش کدگذاری شده بدون زمینه (‏برای مثال، جایگذاری کلمات در berT) ‏و یک سند کدگذاری شده مبتنی بر زمینه (‏برای مثال، کدگذار berT)، ‏ یک طبقه­بندی کننده عصبی را برای محاسبه میزان تطابق نشانه­ها در پرسش و اسناد توسعه می­دهد. | برای معیارهای F1 و EM در مجموعه داده­ OpenSQuAD به ترتیب مقادیر 5/66 و 3/59 و در مجموعه داده OpenNQ مقادیر 8/36 و 5/37 را کسب کرده است. |
| Y. Lin, H. Ji, Z. Liu, and M. Sun | 2018 | TriviaQA | اسناد بازیابی شده را بر اساس احتمال وجود پاسخ رتبه‌بندی می‌کنند و یک محدوده پاسخ را از پاراگراف انتخاب‌شده توسط ماژول انتخاب­گر پاراگراف، استخراج می‌کند. | در مجموعه داده TriviaQA، مقدار 3/56 را برای معیار F1 و 7/48 را برای معیار EM بدست آورده است. |
| S. Min, D. Chen, L. Zettlemoyer, and H. Hajishirzi | 2019 | TriviaQA | ابتدا گراف را به عنوان ورودی گرفته و با استفاده از شبکه‌های کانولوشن گراف بازنمایی­های متن را یاد می‌گیرد و سپس محدوده پاسخ را از محتمل‌ترین محدوده استخراج می‌کند. | روش بر روی مجموعه داده TriviaQA به مقدار 56 درصد برای معیار F1 دست یافته است |
| A. Asai, K. Hashimoto, H. Hajishirzi, R. Socher, and C. Xiong | 2020 | SQuAD | بازیاب مسیر‏ نیز از خواننده برت استفاده می‌کند تا به طور همزمان مسیرهای استدلال را دوباره رتبه‌بندی کند و محدوده پاسخ را از محتمل­ترین کاندیدهای پاسخ، که با استفاده از یادگیری چند وظیفه‌ای بدست آمده، استخراج کند | روی مجموعه داده SQuAD به مقدار 4/72 برای معیار F1 و 5/59 برای معیار EM دست یافته است. |
| C. Tan, F. Wei, N. Yang, B. Du, W. Lv, and M. Zhou | 2018 | MS-MARCO | روش‌های استخراج و تولید را ترکیب می­نماید تا یکدیگر را کامل کنند. این روش از یک مدل استخراج شواهد برای پیش‌بینی مرزهای مهم پاسخ در متون استفاده می‌کند و سپس آنها را به عنوان ورودی به یک مدل سنتز پاسخ پایانه به پایانه می­دهد تا پاسخ نهایی را تولید کند. | در معیارهای کارایی ROUGE-L و BLEU-1 در مجموعه داده MS-MARCO به ترتیب به مقادیر 65/46 و 78/44 دست یافته است. |
| P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. K¨ uttler, M. Lewis, W.-t. Yih, T. Rockt¨aschel, S. Riedel, and D. Kiela | 2020 | Natural Questions,  WebQuestions,  TriviaQA | یک مدل BART از پیش آموزش‌دیده را به عنوان خواننده‌اش در نظر می­گیرد و برای تولید پاسخ، پرسش ورودی و همچنین اسناد بازیابی شده توسط DPR را به عنوان ورودی می­گیرد | معیارهای کارایی EM در مجموعه داده­های NQ، TQA و WQ به ترتیب دارای مقادیر 9/44، 8/55 و 44 می­باشد. |
| G. Izacard and E. Grave | 2020 | ویکی پدیا | هر سند بازیابی شده را به طور مستقل با استفاده از کدگذار T5 یا BART کدگذاری می‌کند و سپس برای تولید پاسخ نهایی، مکانیزم توجه را با استفاده از کدگشا، بر روی تمام بازنمایی‌های خروجی به کار می­گیرد. با این وجود، نتایج بدست آمده کنونی اغلب از خطای نحوی، عدم انسجام یا منطق رنج می‌برند | این روش بر روی مجموعه داده ویکی­پدیا به دقت 1/44 دست یافته است. |
| S. Wang, M. Yu, J. Jiang, W. Zhang, X. Guo, S. Chang, Z. Wang, T. Klinger, G. Tesauro, and M. Campbell | 2018 | Quasar-T, SearchQA, TriviaQA | دو رتبه­بندی کننده پاسخ ، (یک "رتبه­بندی کننده مبتنی بر توان" و یک "رتبه­بندی کننده مبتنی بر پوشش") را برای جمع‌آوری شواهد از متون مختلف به منظور تصمیم‌گیری در مورد پاسخ نهایی پیشنهاد می‌کند. در این رتبه­بند، ابتدا بخش‌هایی را که حاوی پاسخ­های مشابه هستند را در یک شبه متن به هم متصل می‌کند و سپس از شبه متن به دست آمده پاسخ را استخراج می­کند. | در مجموعه داده­های مذکور، معیار EM به ترتیب 3/42، 0/57 و 6/50 و معیار F1 به ترتیب 6/49، 2/63 و 3/57 گزارش شده است. |
| B. Kratzwald, A. Eigenmann, and S. Feuerriegel | 2019 | SQuAD, CuratedTREC, WikiMovies, و WebQuestions | ‏یک ماژول رتبه‌بندی پاسخ را توسعه می‌دهد که شامل سه مرحله است: استخراج ویژگی، تجمیع پاسخ و رتبه‌بندی مجدد. این ماژول نخست، K کاندید پاسخ برتر بدست آمده از خواننده را به عنوان ورودی در نظر گرفته، سپس مجموعه‌ای از ویژگی‌ها را از بازیاب و خواننده استخراج می‌کند. در مرحله دوم، همه کاندیدها در یک محدوده پاسخ یکسان را برای تولید مجموعه‌ای از ویژگی‌های تجمعی مانند جمع، میانگین، حداقل و حداکثر امتیازات محدوده و غیره، گروه­بندی می­کند. در نهایت در مرحله سوم، براساس این ویژگی‌ها، یک شبکه رتبه‌بندی مجدد، مانند یک شبکه پیش‌خور یا یک RNN، برای یادگیری رتبه‌بندی مجدد پاسخ‌ها، برای انتخاب بهترین پاسخ به عنوان پاسخ نهایی مورد استفاده قرار می‌گیرد. | معیار کارایی EM برای چهار مجموعه داده SQuADOPEN، CuratedTREC، WebQuestions و WikiMovies به ترتیب مقادیر 2/54، 9/65، 8/53 و 65 گزارش شده است. |
| M. Seo, J. Lee, T. Kwiatkowski, A. Parikh, A. Farhadi, and H. Hajishirzi | 2019 | SQuAD | با توجه به مجموعه‌ای از اسناد مانند مقالات ویکی‌پدیا، شاخص­گذاری عبارات را به صورت آفلاین ایجاد می‌کند. برای استنتاج، پرسش داده‌شده به همان روش کدگذاری می‌شود FAISS برای جستجوی عبارات مشابه به عنوان پاسخ نهایی به کار گرفته می‌شود. | نتایج این روش در مجموعه داده SQuAD مقادیر 3/42 برای معیار F1 و 4/33 برای معیار EM می­باشد. |
| A. Radford, J.Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever | 2019 | SQuAD | در این روش، مقدار زیادی از دانش آموخته‌شده از داده‌های متنی در مقیاس بزرگ می‌تواند در پارامترهای خاصی ذخیره شوند، و بنابراین این مدل‌ها قادر به پاسخگویی به پرسش­ها بدون نیاز به دسترسی به دانش خارجی هستند. | عملکرد آن بر روی مجموعه داده SQuAD، برای معیار دقت 1/63 درصد گزارش می­شود. |
| A. Roberts, C. Raffel, and N. Shazeer | 2020 | Natural Questions,  WebQuestions,  TriviaQA | به طور جامع قابلیت مدل‌های زبانی برای پاسخ به پرسش­ها، بدون دسترسی به دانش خارجی را ارزیابی کرده است. آزمایش‌ها نشان می‌دهد که مدل‌های زبانی از پیش آموزش‌دیده، قادر به دستیابی به عملکرد چشمگیر در ­benchmarkهای مختلف هستند . | در مجموعه داده­های مذکور به عملکرد 2/35 در مجموعه داده NQ، 8/42 در مجموعه داده WQ و 6/61 در مجموعه داد TQA دست یافته است. |

### 2-8-1- چالش‌های ساخت سیستم­های پرسش و پاسخ

ساخت یک سیستم پرسش و پاسخ که قادر به پاسخگویی به پرسش‌های ورودی باشد، به عنوان هدف نهایی تحقیق پرسش و پاسخ در نظر گرفته می‌شود. با این حال، جامعه تحقیقاتی هنوز راه درازی را در پیش دارد. در اینجا در مورد برخی از چالش‌های برجسته، که باید در این راه به آنها پرداخته شود، بحث می­شود.

### 2-8-1-1- نظارت از راه دور

در محیط سیستم‌های پرسش و پاسخ ، ایجاد یک مجموعه حاوی داده‌های آموزشی "کافی" با کیفیت بالا برای توسعه این سیستم‌ها، غیرممکن است. بنابراین عموماً، نظارت از راه دور مورد استفاده قرار می‌گیرد، که قادر به برچسب زدن خودکار داده‌ها براساس یک پیکره موجود مانند ویکی‌پدیا است. با این حال، نظارت از راه دور به ناچار از مشکل برچسب اشتباه رنج می‌برد و اغلب منجر به حجم قابل‌توجهی از داده‌های نویز می‌شود که دشواری مدل‌سازی و آموزش را افزایش می‌دهد. بنابراین، سیستم‌هایی که قادر به تحمل چنین نویزهایی هستند، همیشه مورد نیاز هستند.

### 2-8-1-2- بازیابی مؤثر و کارایی

کارایی بازیابی به معنای توانایی سیستم در جداسازی اسناد مرتبط از اسناد غیرمرتبط برای یک پرسش مشخص است. سیستم اغلب از "عدم تطابق" رنج می‌برد، که منجر به شکست در بازیابی اسناد مرتبط می‌شود؛ از سوی دیگر، سیستم ممکن است اسناد نویزی را دریافت کند که حاوی اصطلاحات دقیق در پرسش یا حتی محدوده پاسخ صحیح باشد، اما بی‌ارتباط به پرسش باشد. هر دو مساله دشواری درک دقیق مفهوم، برای استنتاج پاسخ را افزایش می‌دهد. برخی از روش‌های بازیابی عصبی [28]‏، [29]‏، [44]‏، [49]‏، [16]‏، [۱72]‏ اخیرا برای بهبود اثربخشی بازیابی پیشنهاد شده‌اند. در این روش­ها ماژول­های بازیاب و خواننده مشترکاً آموزش داده می‌شوند، که از مزیّت­های مدل‌های زبانی از پیش آموزش‌دیده بهره می‌برند و مدل بازیابی را به عنوان یک متغیر پنهان در نظر می‌گیرند.

معمولاً الگوریتم‌های خطی پیشرفته جستجوی ضرب داخلی بیشینه (‏MIPS [[210]](#footnote-210))‏[‏4]‏، [47]‏،‏ برای به دست آوردن K سند برتر مرتبط با یک پرسش بکار گرفته می‌شوند. با این حال، هنگامی که سیستم با مجموعه عظیمی از اسناد مواجه می‌شود، سرعت پاسخ، اختلاف زیادی با تکنیک‌های معمول بازیابی اطلاعات دارد. کارایی و بازیابی مؤثر، هر دو عامل حیاتی برای ساخت یک سیستم پرسش و پاسخ هستند. چگونگی بهبود مستمر هر دو جنبه (‏با یک توازن مناسب) ‏یک چالش قدیمی در پیشرفت سیستم­های پرسش و پاسخ می­باشد.

### 2-8-1-3- تلفیق دانش

ترکیب دانش اسناد (به عنوان مثال، دانش جهانی، دانش مشترک یا دانش دامنه) و پرسش‌های ارائه ‌شده، یک نکته کلیدی برای بهبود سیستم‌های پرسش و پاسخ است [19]‏. قبل از استفاده از چنین دانشی، باید ابتدا نحوه بازنمایی آنها در نظر گرفته شود. به طور کلی دو روش وجود دارد: صریح و ضمنی.

در روش صریح، دانش معمولاً به شکل سه تایی تبدیل می‌شود و در پایگاه­های کلاسیک مانند DBPedia [‏173]‏، فری بیس [174]‏ و Yago2 [۱75]‏ ذخیره می‌شود که به راحتی توسط انسان‌ها درک می‌شود. برخی از سیستم‌های پرسش و پاسخ اولیه تلاش کردند تا به این روش از دانش، برای یافتن پاسخ ، استفاده نمایند. به عنوان مثال، سیستم پرسش و پاسخ عمیق ارائه شده توسط فروسی[[211]](#footnote-211) و همکاران در سال 2010 [181] به نامIBM Watson، یک موتور جستجوی وب و یک پایگاه دانش را برای رقابت با انسان در برنامه تلویزیونی آمریکایی "Jeopardy" ترکیب کرده است؛ و یا سیستم ارائه شده توسط سان[[212]](#footnote-212) و همکاران در سال 2015 [176] به نام QuASE، که لیستی از جملات برتر را از یک موتور جستجوی وب پیدا می­کند ( برای مثال google.com) و سپس با استفاده از پیوند موجودیت­ها، بر مبنای فری بیس [‏۱74]‏ پاسخ صحیح را از جملات انتخاب‌شده تشخیص می‌دهد. در سال‌های اخیر، با محبوبیت شبکه عصبی گراف (‏GNN)‏، برخی کارها مانند سیستم­­های ارائه شده توسط مین[[213]](#footnote-213) و همکاران در سال 2019 [145] و سان و همکاران در سال 2018 [۱77]‏ و [۱78]‏ پیشنهاد می­کنند که اطلاعات مرتبط می­توانند هم از یک پیکره متنی و هم از یک پایگاه دانش بدست آید، که این امر باعث تسهیل در بازیابی شواهد و پاسخگویی به پرسش­ها می­گردد. برای مثال، [۱77]‏ یک زیر گراف مخصوص پرسش، که حاوی جملات بدست آمده از پیکره، نهادها و روابط بدست آمده از پایگاه دانش است، را می­سازد. سپس روش‌های مبتنی بر گراف مبتنی بر CNN [۱60]‏، [۱79]‏، [۱80]‏ برای استنتاج پاسخ نهایی، در زیر گراف ساخته شده، مورد استفاده قرار می‌گیرند. اما در هر صورت، برای ذخیره‌سازی دانش به شیوه‌ای صریح مشکلاتی وجود دارد که عبارتند از: الف)دانش ناقص و ب) به روز نبودن دانش. علاوه بر این، ساخت یک پایگاه دانش هم سخت و هم وقت­گیر است.

از سوی دیگر، با رویکرد ضمنی، حجم زیادی از دانش [۱69]‏ می‌تواند در پارامترهای اساسی آموخته‌شده از متون بزرگ، توسط مدل‌های زبانی از پیش آموزش‌دیده‌ای مانند برت [۲۷]‏، XLNet [‏۱۲۶]‏ و T5 [۱63]‏ ذخیره شود. اخیراً، مدل‌های زبانی از پیش آموزش‌دیده به طور گسترده مورد تحقیق و بررسی قرار گرفته و برای توسعه سیستم‌های پرسش و پاسخ استفاده می­شوند [29]، [44]، [46]، [49]، [132]، [143]، [144]‏. به عنوان مثال، سیستم ارائه شده توسط نی [46] (توضیح داده شده در بخش 2-6-1-1-2‏)، سیستم ارائه شده توسط یانگ [132]‏ (توضیح داده شده در بخش 2-6-1-1-1‏) و سیستم ارائه شده توسط اسائی [144]‏ (توضیح داده شده در بخش2-6‏)، تنها خواننده خود را با استفاده از برت توسعه می‌دهند در حالی که سیستم­های ارائه شده توسط کارپوخین [‏29]‏ (توضیح داده شده در بخش 2-6-1-2) و لی و همکاران [‏49]‏ (توضیح داده شده در بخش 2-5‏) از برت برای توسعه هر دو بخش بازیاب و خواننده استفاده می‌کنند. علاوه بر این، مدل‌های زبان از پیش آموزش‌دیده مانند GPT-2 [۱67]‏ قادر به تولید پاسخ تنها با یک پرسش زبان طبیعی هستند. با این حال، چنین سیستم‌هایی مانند یک "جعبه سیاه[[214]](#footnote-214)" عمل می‌کنند و تقریبا غیر ممکن است که بدانیم چه دانشی دقیقا ذخیره شده‌ و برای پاسخگویی استفاده می­شود. آنها فاقد قابلیت تفسیر هستند که این قابلیت، برای کاربردهای دنیای واقعی بسیار مهم است. ارتقا دانش در سیستم­های پرسش و پاسخ، نه تنها به این دلیل مطلوب است که برای ایجاد پاسخ مفید است بلکه به عنوان منبعی برای تفسیر پاسخ به‌دست‌آمده عمل می‌کند. چگونگی ارائه و استفاده کامل از این دانش برای سیستم­های پرسش و پاسخ هنوز نیازمند تلاش‌های تحقیقاتی بیشتری است.

### 2-9- سیستم­های پرسش و پاسخ محاوره­ای[[215]](#footnote-215)

سیستم­های پرسش و پاسخ غیرمحاوره‌ای با مشکلات متعددی مواجه هستند که حل آنها تقریبا غیرممکن است، برای مثال، وجود کلمات طولانی برای یک پرسش پیچیده ( ‏به عنوان مثال، چه کسی پسر دوم اولین نخست‌وزیر سنگاپور است؟)، وجود ابهام که منجر به پاسخ نادرست می‌شود (‏به عنوان مثال، مایکل جردن چه زمانی متولد شد؟) و یا دانش ناکافی زمینه­ای از کاربر که منجر به نتایج غیر منطقی می‌شود (‏به عنوان مثال، چرا من امروز سردرد بدی دارم؟) این مشکلات تحت محیط محاوره­ای به خوبی قابل حل خواهند بود.

سیستم‌های محاوره‌ای، مجهز به یک رابط گفتگو مانند هستند که تعامل بین کاربران انسانی و سیستم را برای تبادل اطلاعات ممکن می‌سازد [۱82]‏، [۱83]‏. برای مثال پرسش پیچیده‌ای که پیش­تر مطرح شد را می­توان به دو پرسش ساده­تر تجزیه کرد: " اولین نخست‌وزیر سنگاپور چه کسی است؟" و "چه کسی پسر دوم اوست؟". هنگامی که ابهام در پرسش یافت ‌شود، انتظار می‌رود که سیستم پرسش و پاسخ محاوره‌ای یک پرسش پیگیری برای شفاف‌سازی ایجاد کند، مانند " آیا منظور شما بازیکن بسکتبال است؟". اگر پرسشی با دانش پس‌زمینه ناکافی مطرح شود، یک سوال پیگیری نیز می‌تواند برای جمع‌آوری اطلاعات بیشتر از کاربران انسانی برای رسیدن به پاسخ نهایی پرسیده شود. برای رسیدن به این اهداف، سه چالش اصلی باید مورد توجه قرار گیرند. نخست، سیستم پرسش و پاسخ محاوره‌ای باید توانایی تعیین اینکه آیا پرسشی قابل پاسخگویی است یا خیر، را داشته باشد و همچنین توانایی لازم برای تشخیص این که آیا ابهامی در پرسش مطرح شده وجود دارد و یا این که آیا زمینه فعلی برای پاسخگویی به پرسش کافی است، را دارا باشد. تحقیقات در مورد پرسش­های بی­پاسخ، توجه زیادی را در سیستم­های درک مطلب عصبی، در چند سال اخیر به خود معطوف کرده ‌است [120]، [122]، [144]، [184] و [185]. با این وجود، سیستم­های پرسش و پاسخ کنونی به ندرت از چنین مکانیزمی برای تعیین پاسخ ناپذیری پرسش‌ها استفاده می‌کنند.

دوم، هنگامی که پرسش به دلیل ابهام یا دانش پس‌زمینه ناکافی به عنوان پاسخ ناپذیر طبقه‌بندی می‌شود، سیستم پرسش و پاسخ محاوره‌ای نیاز به ایجاد یک پرسش پیگیری دارد [۱86]‏. تولید پرسش (‏QG[[216]](#footnote-216))‏را می‌توان به عنوان بخشی از وظایف مهم از سیستم­های پرسش و پاسخ محاوره‌ای است. در چند سال گذشته، تحقیقات در مورد تولید خودکار پرسش از متن توجه زیادی را به خود جلب کرده‌است [۱87]‏، [۱88]‏، [۱89]‏.

چالش سوم ، چگونگی هر چه بهتر مدل­سازی تاریخچه مکالمه چه در خواننده و چه در بازیاب است [۱91]‏. مجموعه داده‌های ماشین درک مطلب محاوره‌ای مانند CoQA [۱92]‏ و QuAC [‏۱90]‏ خواننده را قادر می­سازد تا برای پاسخ به پرسش، علاوه بر درک متون موجود، از ادراک تاریخچه مکالمات، نیز استفاده نماید. با استفاده از این متون مفهومی، عملاً مرحله بازیابی اسناد را که در مورد سیستم­های پرسش و پاسخ ضروری است، حذف می‌شود. اخیراً، چن­کیو[[217]](#footnote-217) و همکاران در سال 2020 [۱91]‏، یک سیستم بازیابی محاوره­ای (OpenCOQA)، را معرفی کردند، که در آن تلاش می‌شود تا بدون هیچ زمینه خاصی به یک پرسش پاسخ دهد و بنابراین از دامنه کاربرد وسیع­تری برخوردار است و با رفتار پرسش و پاسخ دنیای واقعی انسان سازگاری بهتری دارد. عملکرد این سیستم در مجموعه داده OR - QuAC، (4/29 : F1) بسیار کم‌تر از عملکرد آن در مجموعه داده QuAC (41/74 : F1) است، که نشان می‌دهد این یک چالش بزرگ‌تر در زمینه بازیابی باز است.

### 2-10- مجموعه­های داده­ها

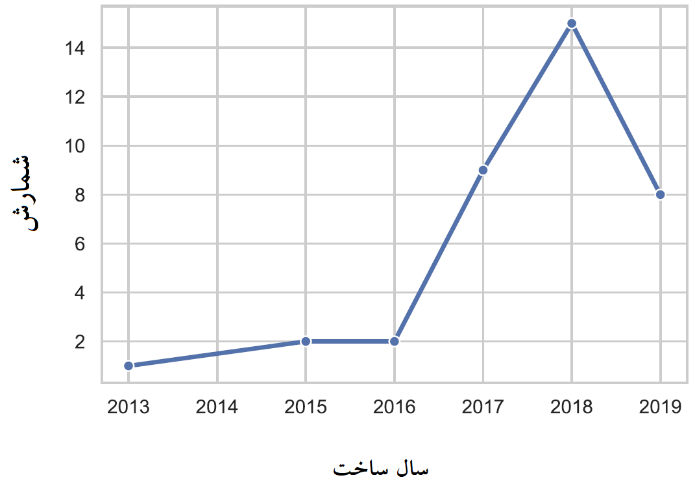
تعداد زیادی از معیارهای پرسش و پاسخ منتشر شده‌اند که در جدول ۲-4 خلاصه شده‌اند. در این قسمت با تمرکز بر ویژگی­های مربوط به آنها، دامنه اطلاعات پس زمینه، توزیع مجموعه داده­ها ، تعداد پرسش­ها و سال انتشار، تجزیه و تحلیل مختصری انجام شده ‌است. همانطور که پیش­تر ذکر شد، موفقیت کار ماشین درک مطلب یک گام مهم برای سیستم­های پرسش و پاسخ پیشرفته‌تر است و پیشرفت آینده روش‌های ماشین­های درک مطلب به طور قابل‌توجهی سیستم‌های پرسش و پاسخ را ارتقا خواهد داد. بنابراین، نه تنها مجموعه داده‌های سیستم­های پرسش و پاسخ، بلکه مجموعه داده‌هایی را که صرفا برای ماشین درک مطلب هستند نیز در نظر گرفته می­شود تا بررسی جامع‌تری انجام شود. معیار اصلی قضاوت در مورد کاربرد یک مجموعه داده پرسش و پاسخ، برای توسعه سیستم‌های پرسش و پاسخ این است که آیا شامل یک مجموعه سند جداگانه (‏معمولا مقیاس بزرگ)‏[۹۰]‏ می­باشد؟ و یا این که آیا دسترسی نسبتا آسانی به چنین منبع اطلاعاتی دارد [119]‏، [1۲۲]‏؟ در سیستم ارائه شده توسط چن

جدول 2-4- انواع مجموعه داده‌هایی که برای توسعه سیستم‌های پرسش و پاسخ قابل بکارگیری هستند

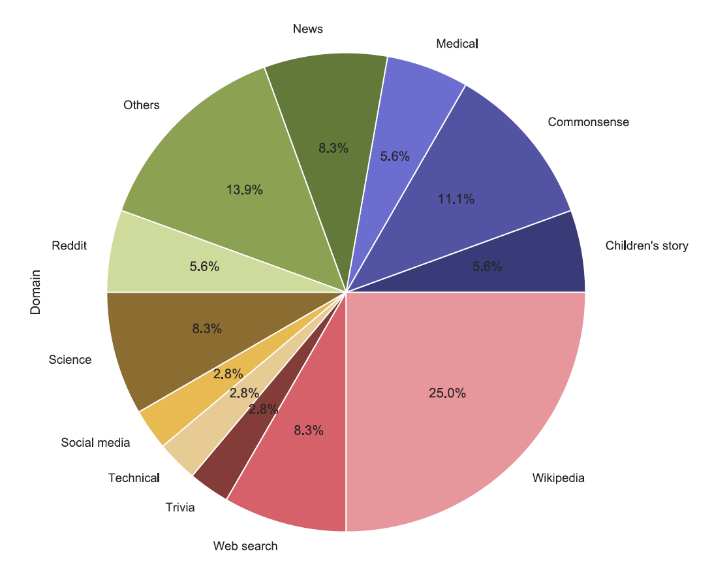
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **مجموعه داده** | **دامنه** | **تعداد پرسش­ها**  **(بر مبنای1000 تا)** | **نوع پاسخ** | **توضیحات** |
| MCTest [102] | داستان بچه­ها | 2 | چند گزینه­ای | داستان­های بچه­گانه |
| CNN/Daily Mail [18] | اخبار | 8/1384 | موجودیت­ها | متونی در مورد اخبار CNN یا Daily Mail |
| CBT [127] | داستان بچه­ها | 3/678 | چند گزینه­ای | داستان­های بچه­گانه |
| SQuAD [19] | ویکی پدیا | 108 | متن | متونی از ویکی پدیا |
| MS MARCO [20] | جستجوی وب | 9/1010 | فرم آزاد | متون مختلف بدست آمده از اینترنت با Bing |
| NewsQA [128] | اخبار | 6/119 | متن | متون اخبار CNN |
| SearchQA [129] | جستجوی وب | 4/140 | متن | متون مختلف بدست آمده از اینترنت با google |
| TriviaQA [130] | Trivia | 9/95 | متن | یک متن یا متن چندگانه |
| RACE [21] | علمی | 6/97 | چند گزینه­ای | متون بدست آمده از امتحانات مدارس |
| Quasar-T [91] | Reddit | 43 | فرم آزاد | اسناد مختلف از Reddit |
| Quasar-S [91] | تکنیکال | 37 | متن علمی | متونی از Stack Overflow |
| **مجموعه داده** | **دامنه** | **تعداد پرسش­ها**  **(بر مبنای1000 تا)** | **نوع پاسخ** | **توضیحات** |
| NarrativeQA [131] | جستجوی وب | 7/46 | فرم آزاد | متون کوتاه و بلند از فیلم­ |
| DuReader [132] | جستجوی وب | 200 | فرم آزاد | متون مختلف از Baidu |
| SQuAD 2.0 [22] | ویکی­پدیا | 158 | متن | متون بدست آمده از ویکی­پدیا |
| CoQA [133] | - | 127 | فرم آزاد | متون و تاریخچه مکالمات |
| QuAC [134] | ویکی­پدیا | 4/98 | متن | متون بدست آمده از ویکی­پدیا و تاریخچه مکالمات |
| ARC [135] | علمی | 7/7 | چندگزینه­ای | - |
| ShARC [136] | - | 4/32 | بولین | تاریخچه مکالمات و سناریوها |
| CliCR [82] | پزشکی | 9/104 | متن | متون مربوط به گزارشات پزشکی |
| HotpotQA [90] | ویکی­پدیا | 113 | متن | پارگراف از ویکی­پدیا |
| MultiRC [138] | - | 6 | چند گزینه­ای | جملات چدتایی |
| SWAG [139] | - | 113 | چند گزینه­ای | متون زیرنویس فیلم­ها |
| DuoRC [140] | - | 186 | فرم آزاد | داستان فیلم­ها |
| **مجموعه داده** | **دامنه** | **تعداد پرسش­ها**  **(بر مبنای1000 تا)** | **نوع پاسخ** | **توضیحات** |
| WikiHop [91] | ویکی­پدیا | 3/51 | چند گزینه­ای | متون مختلف از ویکی­پدیا |
| MedHop [91] | پزشکی | 5/2 | چند گزینه­ای | متون مختلف از MEDLINE |
| ReCoRD [141] | اخبار | 7/120 | چند گزینه­ای | پیام­های CNN و Daily Mail |
| OpenBookQA [5] | علمی | 9/5 | چند گزینه­ای | انواع کتاب­ها |
| CommonsenseQA [142] | - | 2/12 | چند گزینه­ای | - |
| CODAH [143] |  | 8/2 |  | - |
| DROP [103] | ویکی­پدیا | 5/96 | فرم آزاد | متون از ویکی­پدیا |
| Natural Questions [144] | ویکی­پدیا | 323 | - | مقالات از ویکی­پدیا |
| Cosmos QA [145] | - | 6/35 | چند گزینه­ای | متون مختلف |
| BoolQ [146] | ویکی­پدیا | 16 | بولین | مقالات از ویکی­پدیا |
| ELI5 [147] | Reddit | 272 | فرم آزاد | مجموعه­ای از اسناد وب |
| TWEETQA [148] | رسانه اجتماعی | 7/13 | فرم آزاد | توییت­های توییتر |
| **مجموعه داده** | **دامنه** | **تعداد پرسش­ها**  **(بر مبنای1000 تا)** | **نوع پاسخ** | **توضیحات** |
| XQA [149] | ویکی­پدیا | 6/90 | - | متون از ویکی­پدیا |

و همکاران در سال 2017 [‏41]‏ با توسعه مجموعه داده SQuAD [‏۱۹]‏ و استفاده از ویکی‌پدیا به عنوان منبع اطلاعاتی، مجموعه داده SQuADopen معرفی شده است.

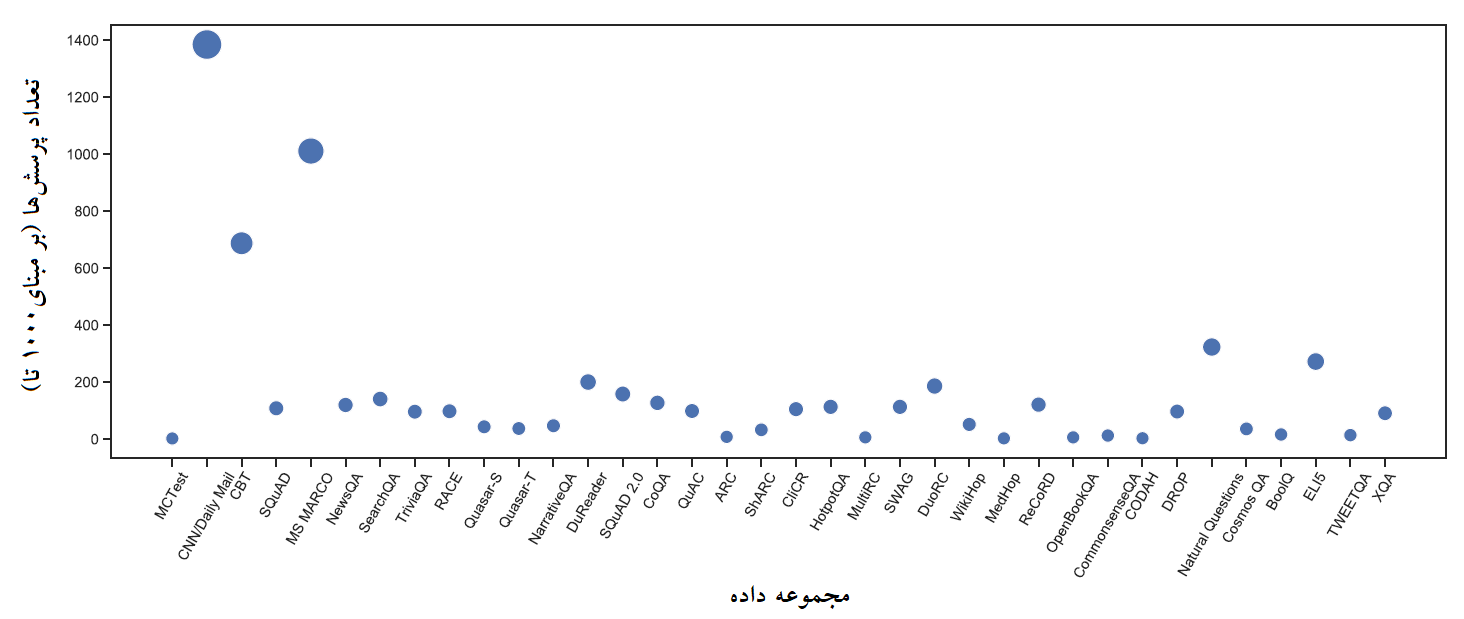
توزیع مجموعه داده­های ذکر شده در جدول 2-4، با توجه به سال انتشار در شکل 2-22، دامنه اطلاعات پس زمینه در شکل 2-23 و تعداد پرسش­ها در شکل 2-24 خلاصه و نشان داده شده است.



شکل 2-22- توزیع مجموعه داده­های ذکر شده، با توجه به سال انتشار



شکل 2-23- دامنه اطلاعات پس زمینه



شکل 2-24- تعداد پرسش­ها در هر مجموعه داده ذکر شده

### 11-2 جمع­بندی

در این فصل مروری بر ادبیات تحقیق در حوزه سیستم­های پرسش و پاسخ انجام شد. انواع دسته­بندی­ها، محدودیت­ها، نقاط قوت­ها و ضعف­ها و چالش­ها در این حوزه بررسی شدند. سه نکته بسیار اساسی و قابل تأمل در این حوزه: نخست، نیاز این سیستم­ها به مجموعه داده­های بزرگ مقیاس، دوم، تعمیم­پذیری کم و نیاز به آموزش مجدد سیستم در زمینه­های جدید و سوم، اصرار به استفاده از یک مدل بسیار قوی که قادر است به تنهایی همه حالت­ها را در نظر گرفته و پاسخ صحیح را تولید کند، می­باشد. حال آنکه در بسیاری از حوزه­ها امکان ایجاد مجموعه داده بزرگ میسر نیست و لذا نیاز به سیستم­هایی است که بتواند با تعداد داده کم و افزایش تعمیم­پذیری نتایج قابل قبولی تولید نماید. از طرف دیگر می­توان به جای استفاده از یک مدل بسیار قوی که نیاز به آموزش فراوانی دارد، از چندین مدل به صورت موازی استفاده شود و هر یک از مدل­های کوچک­تر، تنها در یک حوزه مشخص متخصص گردد و مسئول پاسخگویی به پرسش­های حوزه تخصصی خود باشند.

فصل 3

روش پیشنهادی

**1-3- مقدمه**

سیستم‌های پرسش و پاسخ زیر شاخه ای از علوم پردازش زبان طبیعی و بازیابی اطلاعات محسوب می‌شوند. سیستم‌های پرسش و پاسخ، پرسش را به یک زبان طبیعی (مثلاً فارسی) دریافت کرده و جواب مختصر و دقیق را در اختیار کاربر قرار می‌دهند. بنابراین دیگر لازم نیست کاربر مانند سیستم‌های بازیابی اطلاعات پرسش خود را به کلید واژه‌‌ها تبدیل کند و پس از بازیابی، تعداد زیادی سند را مطالعه کند تا به جواب دلخواه خود برسد.

در فصل پیش انواع سیستم­های پرسش و پاسخ و نحوه عملکرد آنها مورد بررسی قرار گرفت. تفکّر غالب بر سیستم­های پرسش، استفاده از یک "کلید طلایی" برای حل تمامی مسائل می­باشد. منظور از کلید طلایی استفاده از یک راه حل جامع برای پاسخگویی به تمامی پرسش­ها می­باشد. لذا مهمترین مانعی که در سیستم­های پرسش و پاسخ مطرح شده، قابل مشاهده است، عدم ثبات در کارآیی در مجموعه داده­های مختلف می­باشد. ممکن است روشی پیشنهاد شود که در یک مجموعه داده کارآیی مناسبی داشته باشد. اما در مجموعه­های دیگر کارآیی مناسبی نداشته و می­بایست مجدّدا در اکثر موارد از صفر آموزش داده شوند. البته روش­هایی مانند انتقال یادگیری مطرح شده است که فرآیند آموزش برای مجموعه داده جدید را تسریع می­سازد. در این فصل یک سیستم پرسش و پاسخ پایانه به پایانه رقابتی ماژولار[[218]](#footnote-218)، پیشنهاد می­گردد که در آن به پرسش جزئی­تر و دقیق­تر پرداخته شده و ماژول­ها در ارائه پاسخ با یکدیگر به رقابت پرداخته و در هر مرحله بهترین پاسخ به عنوان پاسخ منتخب برگزیده می­گردد. این امر سبب می­گردد تا هر ماژول در ارائه پاسخ به حوزه مشخص و متفاوتی متخصص گردد. این روش قادر است،­ علاوه بر ارائه پاسخ­های مناسب و کارآیی بالا، مشکلات مذکور را نیز برطرف سازد.

**2-3- روش پیشنهادی**

معماری روش پیشنهادی بر سه بخش اساسی لایه ورودی، متخصصین و جداکننده استوار است که در شکل 3-1 قابل مشاهده است.

**پرسش**

**پاسخ**

**ورودی**

E1

EN

E2

E3

**متخصصین**

**پاسخ1**

**پاسخ2**

**پاسخN**

**پاسخ3**

**پاسخ صحیح**

**جدا** **کننده**

0.2

**0.5**

0.1

0.0

**جدا کننده**

**argmax**

شکل 3-1- معماری کلی روش پیشنهادی

* **لایه ورودی**، در این لایه جفت­های پرسش و پاسخ پیش پردازش شده و به صورت توکن­های مجزا به مرحله متخصصین اعمال می­شوند.
* **لایه متخصصین**، از تعدادی متخصص تشکیل شده که هریک از آنها در یک فرآیند رقابتی برای پاسخگویی به پرسش­های حوزه خاصی آموزش خواهند دید.
* **لایه جداکننده**، وظیفه مقایسه پاسخ­های بدست آمده از متخصصین و انتخاب بهترین پاسخ و درنتیجه انتخاب بهترین متخصص برای پاسخگویی به پرسش ورودی را دارد.

در ادامه هر یک از لایه­ها مفصلاً تشریح خواهند شد.

**2-3-1- لایه ورودی­**

از آنجایی که این روش به صورت ماژولار و رقابتی است، لذا نیاز به ارائه مجموعه داده­های آموزشی است که شامل جفت­های پرسش و پاسخ بوده تا سیستم قابلیت مقایسه پاسخ تولیدی خود با پاسخ صحیح را در هر مرحله داشته باشد. جفت­های پرسش و پاسخ بعد از توکن­گذاری با بازنمایی مناسب در اختیار مرحله بعد قرار می­گیرد.

**2-3-2-** **لایه متخصصین**

لایه دوم، لایه متخصصین است که قسمت اصلی الگوریتم پیشنهادی می­باشند. در ادامه چگونگی کارکرد این بخش مورد بررسی قرار می­گیرد.

در روش پیشنهادی، از مجموعه متخصصین برای تولید پاسخ استفاده می­شود. در حقیقت این رویکرد بر مجموعه­ای از توابع متخصص متکی است که برای تولید پاسخ، طی مکانیزم­های مختلف برای کسب تخصص به رقابت می­پردازند که این مکانیزم­ها ماژولار بوده و قابلیت استفاده­ی مجدد و وسیع دارند.

برای سهولت در بحث، ابتدا عملکرد یک متخصص به تنهایی مورد بررسی قرار گرفته و سپس به بحث در مورد نحوه تعامل متخصصین به منظور کسب مهارت و تخصص در پاسخگویی به پرسش­ها، پرداخته خواهد شد.

انتخاب مدل برای متخصصین یکی از نکات چاش برانگیز در روش پیشنهادی است. دو رویکرد برای این امر وجود دارد، که عبارتند از:

الف) استفاده از مدل­های متفاوت برا ی متخصصین

در این حالت به تعداد متخصصین مدل­های مختلف وجود خواهد داشت که با یکدیگر برای کسب تخصص به رقابت می­پردازند

ب) استفاده از مدل یکسان برای متخصصین

در این حالت یک مدل انتخاب شده و از آن برای همه متخصصین استفاده خواهد شد. از آنجایی که مدل­ها مقداردهی­های اولیه متفاوتی خواهند داشت و در فرآیند آموزش برای پاسخگویی به پرسش­ها رقابت خواهند کرد. به تدریج هر یک در حوزه خاصی متخصص خواهند شد.

چنانچه

N’: تعداد متخصصین ،q’: پرسش ورودی از مجموعه داده DQ ، E: متخصصین

آنگاه پاسخ متخصص jام (Ej) به پرسش q’ که با Cj نشان داده می­شود، از رابطه 3-1 بدست می­آید.

|  |  |
| --- | --- |
| رابطه 3-1 | cj = c (Ej(q’)) |

متخصص با بهترین پاسخ (بیشترین میزان شباهت به پاسخ ایده­آل) نیز از رابطه 3-2 بدست می­آید.

|  |  |
| --- | --- |
| رابطه 3-2 | j\* = arg maxj (cj) |

پس از انتخاب بهترین پاسخ، متخصص منتخب که پارامترهای آن با نشان داده می­شود، آموزش داده شده و بروزرسانی می­شود. در این حالت سایر توابع متخصص بدون تغییر باقی می­مانند. انگیزه از ایجاد اینگونه رقابت و تنها به روز کردن تابع متخصص برنده، این است که بتوان مجموعه ای از توابع متخصص را بر روی مجموعه داده دلخواه آموزش داد. از آنجایی که پرسش­ها در حوزه های متفاوتی مطرح می­شوند لذا هر متخصص پس از مدتی در حوزه خاصی آموزش دیده و پاسخ های بهتری در حوزه به خصوص خود تولید می­نماید.

در نهایت مسأله­ی بهینه­سازی به صورت رابطه 3-3 خواهد بود.

|  |  |
| --- | --- |
| رابطه 3-3 |  |

**2-3-3-** **لایه جداکننده**

بخش جداکننده وظیفه دارد پاسخ­های تولید شده توسط متخصصین را در هر تکرار دریافت کرده و بهترین پاسخ را انتخاب نماید. در ادامه متخصصی که بهترین پاسخ را داشته برای آموزش انتخاب می­شود و دیگر متخصصین بدون هیچ آموزشی وارد تکرار بعدی می­شوند. این نوع نگرش در رقابت بین متخصصین و نحوه آموزش مذکور، منجر به تشکیل متخصصینی در حوزه­های مختلف می­گردد.

با توجه به مطالب بیان شده به صورت خلاصه می­توان گفت ایده اصلی روش پیشنهادی این است که از آنجایی که در پرسش­های مطرح شده زمینه­های متفاوتی وجود دارد. لذا می­توان تعدادی متخصص تعیین نمود که هر یک ازآنها در زمینه خاصی متخصص شده و به پرسش­های حوزه تخصصی خود پاسخ بگوید. در حقیقت مقصود این است که پرسش­ تخصصی­تر دیده شود و به جای استفاده از یک مدل برای پاسخگویی به پرسش­های همه زمینه­ها از چندین مدل (متخصص) استفاده گردد. برای این منظور می­توان حالت­های مختلفی را در نظر گرفت. براین اساس مشارکت این رساله به 3 روش زیر صورت خواهد گرفت.

الف) برای هریک از متخصصین یکی از سیستم­هایی که در فصل پیش مورد بررسی قرار گرفته­اند، در نظر گرفته شود. برای مثال، DrQA، ArQA، T5 و ... . سپس این مدلها به صورت مستقل در یک فرآیند آموزشی رقابتی برای کسب تخصص آموزش ببینند. معماری این روش مطابق با شکل 3-1 می­باشد.

ب) برای هریک از متخصصین یکی از سیستم­هایی که در فصل پیش مورد بررسی قرار گرفته­اند، در نظر گرفته شود. برای مثال، DrQA، ArQA، T5 و ... . سپس این مدلها تبادل اطلاعات داشته باشند. یعنی با به اشتراک گذاشتن دانش خود با دیگر متخصصین که از طریق اشتراک گذاری وزن­ها صورت می­گیرد در فرآیند آموزشی رقابتی برای کسب تخصص آموزش ببینند. معماری این روش در شکل 3-8 قابل مشاهده است.

**ورودی**

**پرسش**

**پاسخ**

E1

**پاسخ1**

**ورودی**

**0.5**

0.2

**پاسخ صحیح**

**جدا کننده**

**متخصصین**

E1

E2

E3

EN

**پاسخ1**

**پاسخ2**

**پاسخN**

**پاسخ3**

**پاسخ صحیح**

**جدا** **کننده**

**جدا کننده**

0.2

**0.5**

0.1

0.0

**argmax**

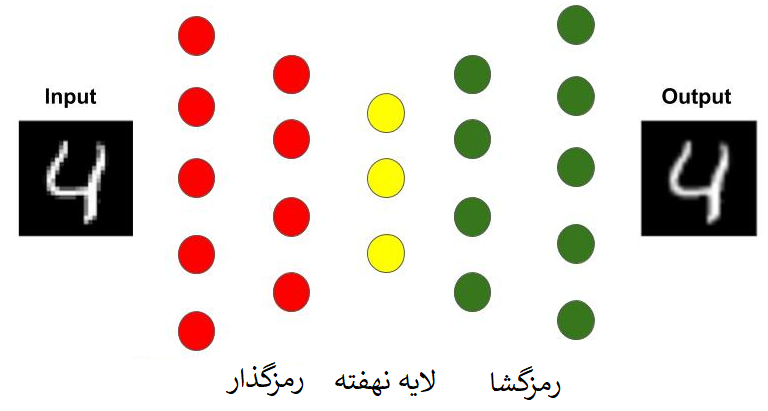
شکل 3-8- معماری روش یادگیری رقابتی متخصصین با اشتراک وزنی

ج) از مفهوم رمزگذار خودکار متغیر برای ارائه روشی جهت پاسخگویی به پرسش­ها استفاده شود. برای بیان روش مورد نظر ابتدا به صورت خلاصه رمزگذار خودکار متغیر معرفی می­شود.

رمزگذارهای خودکار شبکه­های ساده‌ی یادگیری هستند که با هدف تبدیل ورودی به خروجی بدون کمترین تغییر ‌پیاده‌سازی می‌شوند. در عین سادگی رمزگذارهای خودکار نقش بسزایی را در یادگیری ماشین ایفا می‌کنند. برای اولین بار این مفاهیم در سال ‎1980‎ توسط ‎هینتون[[219]](#footnote-219)‎ و گروه تحقیقاتی ‎PDP‎ مطرح شد. رمزگذار خودکار در کنار قوانین یادگیری هِب، از پارادایم‌های اصلی یادگیری بدون نظارت می­باشند. رمزگذار خودکار مجدداً در دهه‌ی ابتدایی قرن بیستم در معماری عمیق به فرم ماشین محدود بولتزمن ‎ مورد توجه قرار گرفتند.

رمزگذار خودکار، یک شبکه‌ی عصبی است که مجموعه‌ای از داده‌های بدون برچسب را دریافت می‌کند و با رمزکردن آن‌ها ‎(encoding)‎ سعی در بازنمایی مجدد ورودی‌ها در خروجی، به صورتی که کمترین اختلاف ممکن را با مقدار ورودی­ داشته باشد، دارد.

شکل 3-2 یک شبکه رمزگذار خودکار را نشان می­دهد. همان­طور که می­بینید شبکه در جهتی آموزش می­بیند که وزن­های تولید شده در لایه­ها، باعث شود خروجی با ورودی حداقل اختلاف ممکن را داشته باشد و در ایده­آل­ترین حالت ممکن برابر شوند.



شکل 3-2 نمونه­ای از یک شبکه رمزگذار خودکار

ساختار رمزگذارهای خودکار به دو بخش رمزگذاری و رمزگشایی ‎تقسیم می‌شود. در بخش رمزگذاری داده‌های ورودی به فضای ویژگی‌ها (لایه نهفته) نگاشت می‌شوند و در بخش رمزگشایی از فضای ویژگی مجدداً به حالت ابتدایی خود تبدیل می‌شوند. در واقع بخش اصلی یک خود رمزگذار لایه‌ی پنهان میانی است که به عنوان ویژگی استخراج شده برای دسته‌بندی استفاده می‌شود.

به عبارت دیگر، رمزگذارهای خودکار، داده‌ها را به عنوان ورودی دریافت می‌کنند و برخی از نمایش‌های حالت پنهان آن داده را کشف می‌کنند. داده­های ورودی به یک بردار رمزگذاری (بازنمایی) با ابعاد مشخص تبدیل می­شوند که در آن هر بعد نشان دهنده برخی ویژگی­های آموخته شده در مورد داده­ها است. شبکه رمزگذار یک مقدار واحد را برای هر بعد به عنوان خروجی بخش رمزگذاری تولید می­نماید. سپس شبکه رمزگشا این مقادیر (بازنمایی) را دریافت نموده و سعی می­نماید تا ورودی اصلی را دوباره ایجاد کند. در واقع در رمزگذارهای خودکار بجای آموزش شبکه و پیش­بینی مقدار هدف Y در ازای ورودی X، آموزش می­بیند تا ورودی X را دوباره بازسازی نماید. بنابراین بردارهای خروجی، همان ابعاد بردار ورودی را خواهند داشت. فرآیند کلی یک رمزگذار خودکار در شکل 3-3 نشان داده شده است. در حین این فرآیند رمزگذار خودکار با کمینه­سازی خطای نوسازی[[220]](#footnote-220) بهینه می­شود.

ورودی

بازنمایی

بازسازی

رمزگذار

رمزگشا

شکل 3-3- معماری رمزگذار خودکار

رمزگذار خودکار متغیر مدل خاصی از رمزگذار خودکار است که در آن از یک روش احتمالی برای توصیف داده ورودی در فضای نهفته استفاده می­شود. بنابراین، در رمزگذار خودکار متغیر، به جای ساخت یک رمزگذار که یک مقدار واحد را برای توصیف ویژگی­های حالت نهفته تولید ­کند، از رمزگذاری استفاده می­شود که برای توصیف ویژگی­های حالت نهفته، از توزیع احتمال استفاده می­نماید.

به عنوان مثال، در شکل 3-4، یک مدل رمزگذار خودکار بر مجموعه داده بزرگی از چهره‌ها با ابعاد کدگذاری 6 آموزش داده شده است. خروجی رمزگذار خودکار شامل 6 ویژگی‌ توصیفی چهره‌ مانند لبخند، رنگ پوست، جنسیت، ریش داشتن یا نداشتن، عینک زدن یا عدم استفاده از عینک و رنگ مو می­باشد.

|  |  |  |
| --- | --- | --- |
| C:\Users\ASUS\Desktop\44607.jpg |  | C:\Users\ASUS\Desktop\44607.jpg |

شکل 3-4- ویژگی­های نهفته تصویر ورودی در یک رمزگذار خودکار

در این شکل، تصویر ورودی بر حسب ویژگی‌های پنهان آن (با ابعاد 6) و با استفاده از یک مقدار واحد برای توصیف هر ویژگی، توصیف شده است.

لبخند: 99/0، رنگ پوست: 85/0، جنسیت: 73/0-، ریش داشتن یا نداشتن: 85/0، عینک زدن یا عدم استفاده از عینک: 002/0 و رنگ مو: 68/0.

همان­گونه که در اَشکال 3-5 و 3-6 قابل مشاهده است، چنانچه هر ویژگی پنهان به جای یک مقدار واحد، به عنوان یک محدوده از مقادیر ممکن نشان داده شود، آنگاه می­توان ویژگی‌های نهفته را با شرایط احتمالی توصیف نمود. در نتیجه، هر ویژگی پنهان مرتبط با یک ورودی معین، به صورت یک توزیع احتمال نشان داده خواهد شد. این امر سبب می­گردد تا در رمزگشا به جای دریافت مقادیر ثابت برای ویژگی­ها، طیفی از مقادیر دریافت شده و با اِعمال آن به رمزگشا، خروجی­های مختلفی تولید شود.

|  |  |  |
| --- | --- | --- |
| عکس­ها | لبخند (توزیع احتمال) | لبخند (مقدار گسسته) |
| C:\Users\ASUS\Desktop\Sad-photo-pose-19.jpg |  | |
| C:\Users\ASUS\Desktop\download (1).jpg |  | |
| C:\Users\ASUS\Desktop\nody-عکس-بچه-کوچولو-بامزه-1626652954.jpg |  | |
| C:\Users\ASUS\Desktop\44607.jpg |  | |

شکل 3-5- استفاده از تابع توزیع احتمال به جای یک مقدار واحد (گسسته) برای ویژگی­ها

|  |  |  |
| --- | --- | --- |
| C:\Users\ASUS\Desktop\44607.jpg |  | C:\Users\ASUS\Desktop\44607.jpg |

شکل 3-6- استفاده از توزیع احتمال برای ویژگی­ها

در حقیقت در حالتی که خروجی رمزگذار ویژگی­های با مقادیر ثابت است، مدل یک رمزگذار خودکار بوده و در حالتی که خروجی رمزگذار ویژگی­های با مقادیر توزیع احتمال است مدل یک رمزگذار خودکار متغیر می­باشد.

هنگام رمزگشایی از حالت نهفته، از هر توزیع حالت نهفته به صورت تصادفی نمونه­برداری می­شود تا یک بردار به عنوان ورودی برای رمزگشا تولید شود. به همین ترتیب همان­گونه که در شکل 3-7 مشاهده می­شود رمزگشا به ازای نمونه برداری­های مختلف از حالت نهفته خروجی­های مختلفی تولید خواهد نمود.

|  |  |  |
| --- | --- | --- |
| C:\Users\ASUS\Desktop\44607.jpg |  | |
| C:\Users\ASUS\Desktop\44607.jpg |
|  | ویژگی­های نمونه­برداری شده | توزیع­های لایه نهفته |

شکل 3-7- خروجی­های مختلف رمزگشا به ازای نمونه برداری­های مختلف از حالت نهفته

با توجه به مطالب بیان شده استفاده از رمزگذار خودکار متغیر باعث تولید انواع داده جدید خواهد شد. لذا استفاده از این مدل برای آموزش متخصصین سبب می­گردد تا هر متخصص قادر باشد پاسخی برای پرسش ورودی تولید نماید. سپس پاسخ­ها با پاسخ ایده­آل مقایسه شده و بهترین پاسخ برگزیده می­شود. در حقیقت تمامی متخصصین برای تولید پاسخ با هم رقابت می­کنند و همگی آنها پاسخی برای پرسش ورودی تولید می­نمایند. سپس پاسخ­ها توسط جدا کننده با پاسخ اصلی مقایسه شده و بهترین پاسخ برگزیده شده و متخصصی که آن پاسخ را تولید نموده برای آموزش انتخاب می­شود.

**پاسخ**

**پرسش**

**ورودی**

**رمزگذار**

**لایه نهفته**

μn

nσ

μ3

3σ

μ2

2σ

μ1

1σ

**رمزگشا (متخصصین)**

EN

E3

E2

E1

**پاسخN**

**پاسخ3**

**پاسخ1**

**پاسخ2**

**پاسخ صحیح**

**جدا** **کننده**

0.01

0.2

**0.5**

0.1

**جدا کننده**

**argmax**

شکل 3-9- معماری روش چیدمان متخصصین در ساختار رمزگذار خودکار متغیر

در روش سوم متخصصین در یک ساختار به نام رمزگذار خودکار متغیر چیده می­شوند. در این روش بر خلاف دو روش پیشین به تعداد متخصصین، مدل وجود ندارد و تنها یک مدل رمزگذار خودکار متغیر وجود دارد که متخصصین در آن چیده می­شوند. همان­گونه که پیش­تر در خصوص معماری رمزگذار خودکار متغیر بیان شد. این مدل از دو بخش رمزگذار و رمزگشا تشکیل شده و بخش رمزگذار وظیفه تعریف توزیع­های احتمال مختلف را برای ورودی مدل (پرسش ورودی) به عهده دارد. می­توان به تعداد n توزیع احتمال مختلف را در نظر گرفت که هر یک از آنها به عنوان یک متخصص تولید پاسخ می­نمایند. این روش همانند روش دوم، در درون خود بحث اشترک گذاری وزن­ها را نیز پوشش می­دهد. معماری این روش در شکل 3-9 قابل مشاهده است.

**3-3-** **انتخاب تعداد متخصصین**

E3

E2

E1

**پاسخ3**

**پاسخ2**

**پاسخ1**

**پرسش**

**ورودی**

0.1

**0.5**

0.2

**argmax**

**پاسخ صحیح**

**جدا کننده**

نکته حائز اهمیت در روش­های پیشنهادی، انتخاب تعداد متخصصین است. سناریوهای مختلفی را برای تعداد متخصصین می­توان در نظر گرفت که در ذیل به آن پرداخته شده است.

عمدتاً تعداد زمینه­های پرسش N که در مجموعه داده­ی DQ وجود دارد در ابتدا مشخص نیست. بنابراین نیاز به یک رویه­ انطباقی برای انتخاب تعداد متخصصین N’ می­باشد. این یکی از چالش­هایی است که در اغلب تکنیک­های خوشه­بندی مشاهده می­شود. با توجه به رفتار ماژولار فرآیند، توابع متخصص را می­توان طی فاز آموزش یا پس از آن اضافه یا کم کرد، و بدین ترتیب چارچوبی بسیار منعطف ایجاد می­شود. با فرض ثابت ماندن تعداد توابع متخصص، رفتار زیر بروز پیدا می­کند:

1- اگر N’>N (توابع متخصص بیش از حد): الف) برخی از این توابع تخصصی­سازی نشده و نمی­تواند در رقابت بر سر هیچ نمونه­ای را در مجموعه داده پیروز شوند، و یا ب) برخی وظایف میان توابع متخصص تقسیم می­شود (مثلاً هر تابع متخصص را می­توان در حالتی از توزیع برای یک وظیفه­ی مشخص تخصصی­سازی کرد). در حالت الف توابع متخصص غیرفعال قابل کنار گذاشته شدن هستند، و در حالت ب توابع متخصصی که وظیفه­ی مشابهی را به اشتراک می­گذارند با هم تلفیق شده و یک تابع بزرگتر را ایجاد می­کنند.

2- اگر N’=N (تعداد توابع برابر با زمینه­های پرسش): این حالت در واقع یک حالت بهینه است. هر یک از توابع در یک زمینه مشخص مهارت پیدا کرده و مسئول پاسخگویی به پرسش­های همان زمینه است.

2- اگر N<N’ (تعداد کمتر از توابع متخصص): الف) برخی از توابع متخصص برای انجام چندین وظیفه تخصصی­سازی می­شوند، و یا ب) برخی از وظایف را نمی­توان به توابع آموزش داد، بنابراین زمینه­های مربوط به این وظایف نتایج ضعیفی را در میان تمامی توابع متخصص بدست می­آورند.

**4-3- معیار همگرایی**

از آنجایی که مسئله غیرنظارت شده است، هیچ راه مستقیم و ساده­ای برای سنجش همگرایی در آن وجود ندارد. بدین ترتیب این پرسش ایجاد می­شود که چگونه می­توان زمان توقفی برای وضعیت رقابتی تعیین کرد. بعنوان مثال می­توان به یکی از این روش­ها عمل کرد: 1- ثابت نگاه داشتن تعداد حداکثر تکرارها، یا 2- توقف در صورتی که هر پرسش به توابع متخصص مشابهی در تعداد از پیش تعیین شده­ای از تکرارها تخصیص داده شود (یعنی هر تابع متخصص بصورت مداوم رقابت بر سر پرسش مشابهی را پیروز شود). 3- ارضا شدن حد آستانه برای یکی از معیارهای ارزیابی مانند دقت، F1، EM، ROUGE-L و BLUE.

**5-3- پیچیدگی زمانی و فضایی**

هر پرسش را باید توسط تمامی توابع متخصص ارزیابی کرد تا امکان تخصیص آن به تابع متخصص برنده وجود داشته باشد. از آنجایی که این مسئله منجر به افزایش هزینه­های محاسباتی متکی بر غیر خطی بودن تعداد توابع متخصص می­شود، چنین ارزیابی­هایی را می­توان به شکل موازی انجام داده و پیچیدگی زمانی یک دوره­ی تکرار را محدود کرد به پیچیدگی محاسبه­ی خروجی یک تابع متخصص. علاوه بر این به دلیل اینکه هر تابع متخصص در اصل دارای معماری کوچکتری از یک شبکه­ی بزرگ منفرد است، جمعی از این توابع می­توانند اجرای سریع­تری داشته باشند.

**6-3- نوآوری پژوهش**

سیستم­های پرسش پاسخ دارای کاربردهای فراوانی در حوزه­های مختلف از جمله نظامی، آموزشی، تحقیقاتی و ... می­باشند. استفاده از یک مدل بزرگ و پیچیده برای پاسخگویی به پرسش­ها منجر به نتایج نه چندان رضایت بخشی خواهد شد. بنابراین در این مستند سه معماری متفاوت برای سیستم­های پرسش و پاسخ پیشنهاد می­شود که هر سه آنها بر پایه ترکیب مدل­های ساده­تر و آموزش رقابتی آنها می­باشد. پیشنهاد نخست، استفاده از مدل­ها متفاوت و مستقل پرسش و پاسخ به گونه­ای که در یک فرآیند آموزشی رقابتی برای کسب تخصص آموزش ببینند. معماری این روش در شکل 3-1 قابل مشاهده است. در پیشنهاد دوم، از مدل­های متفاوت پرسش و پاسخ استفاده می­شود. این مدل­ها مستقل نبوده و با یکدیگر تبادل اطلاعات دارند. یعنی با به اشتراک گذاشتن دانش خود با دیگر متخصصین که از طریق اشتراک گذاری وزن­ها صورت می­گیرد در فرآیند آموزشی رقابتی برای کسب تخصص آموزش می­بینند. معماری این روش در شکل 3-8 قابل مشاهده است. نهایتا در پیشنهاد سوم، متخصصین در یک ساختار به نام رمزگذار خودکار متغیر چیده می­شوند. در این روش بر خلاف دو روش پیشین به تعداد متخصصین، مدل وجود ندارد و تنها یک مدل رمزگذار خودکار متغیر وجود دارد که متخصصین در آن چیده می­شوند. در واقع این روش هر یک از توزیع­های به دست آمده در لایه پنهان رمزگذار خودکار متغیر را به عنوان یک متخصص تولید پاسخ در نظر می­گیرد. معماری این روش در شکل 3-9 قابل مشاهده است.

7**-3- مجموعه داده­های مورد استفاده**

* SQUAD

این مجموعه داده توسط محققان دانشگاه استنفورد ارائه شده است و حاوی بیش از 100000 سوال می-باشد. SQUAD یک مجموعه داده درک مطلب است که شامل سوالاتی است که توسط جمعی از کارگران در مجموعه‌ای از مقالات ویکی‌پدیا مطرح می‌شود و پاسخ به هر سؤال می­تواند بخش یا گستره‌ای از متن مربوط به سؤال و یا ممکن است بی­پاسخ باشد.

* Natural Question(NQ)

سوالات طبیعی (NQ) مجموعه­ای جدید و در مقیاس بزرگ برای آموزش و ارزیابی سیستم های پاسخگویی به سؤالات دامنه باز است که توسط گوگل ارائه شده است. این مجموعه شامل 300000 سؤال طبیعی ، به همراه پاسخ‌های مشروح توسط انسان از صفحات ویکی‌پدیا می­باشد. علاوه بر این، محققان 16000 مثال اضافه کردند که در آن پاسخ‌ها (به سؤالات یکسان) توسط 5 پاسخ دهنده مختلف انسانی ارائه شده است که برای ارزیابی عملکرد سیستم‌های پرسش و پاسخ آموخته شده مفید خواهد بود.

* PersianQA

یک مجموعه داده برای پاسخگویی به پرسش­های فارسی با استفاده از مجموعه­داده­های درک مطلب در ویکی پدیای فارسی است. مجموعه­داده جمع آوری شده از بیش از 9000 ورودی تشکیل شده است. هر ورودی می‌تواند یک پرسش بدون پاسخ یا پرسشی با یک یا چند پاسخ (بر اساس زمینه) که پرسش­گر از آن پرسش را مطرح کرده است.

**8-3- معيارهای ارزیابی**

در این مستند از دو معیار EM و F1 برای ارزیابی و مقایسه روش پیشنهادی با سایر روش­ها استفاده می­شود که در فصل 2 به آنها پرداخته شده است.

معیار EM حالت خاصی از معیار دقت است که در آن، چناچه پاسخ سیستم برابر با پاسخ حقیقی بود مقدار 1 و در غیر این صورت 0 برگردانده می­شود.

معیار F1 بر اساس ماتریس اغتشاش عمل می­نماید. نمونه­ای از ماتریس اغتشاش در شکل 2-8 قابل رویت می­باشد.

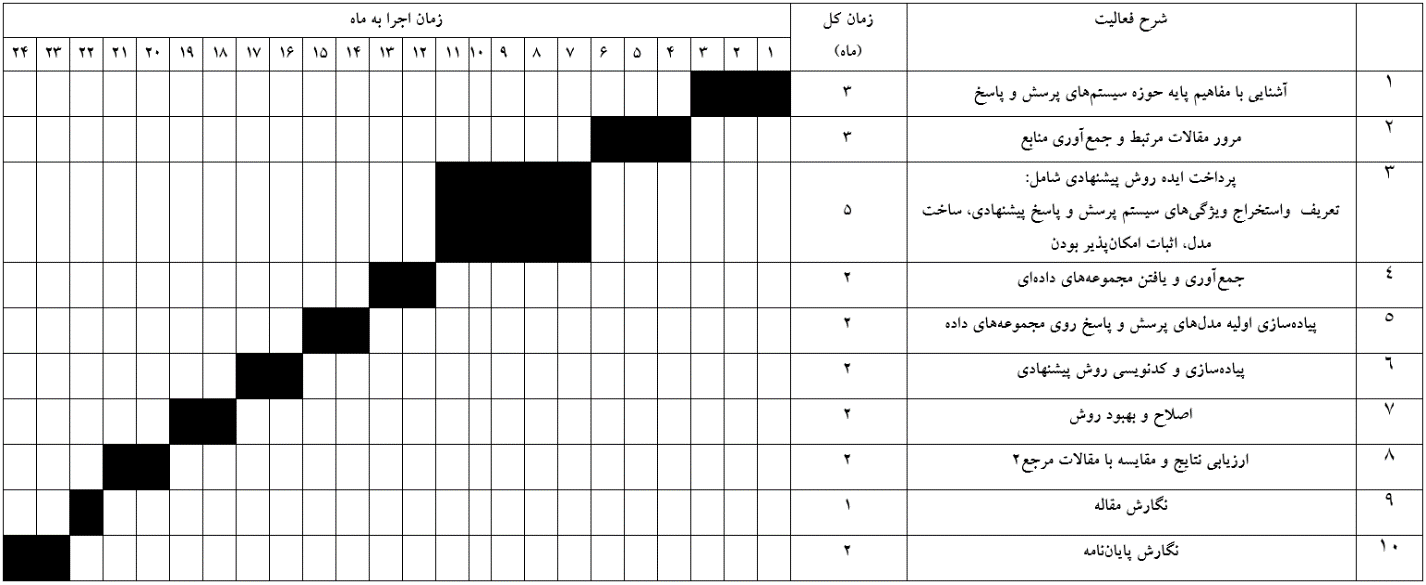
معیار F1 از رابطه 2-5 بدست می­آید.

|  |  |
| --- | --- |
| رابطه 3-4 |  |

که در این رابطه مقادیر P و R به ترتیب از روابط 2-6 و 2-7 محاسبه می­شوند.

|  |  |
| --- | --- |
| رابطه 3-5 |  |
| رابطه 3-6 |  |

**9-3- پيش‏بيني زمان‏بندي فعاليت‏ها و مراحل اجرايي تحقيق و ارائه گزارش پيشرفت كار**



**10-3- جمع­بندی**

سیستم­های پرسش و پاسخی که تاکنون پیشنهاد شده­اند اصرار بر استفاده از یک مدل بسیار قوی که قادر باشد به همه پرسش­های مطرح شده پاسخ مناسبی ارائه دهد، دارند. این تفکر منجر به بوجود آمدن سیستمهایی شده است که غالبا نتوانسته انتظارات را آنگونه که باید برطرف سازد. در این فصل به ارائه سه روش تولید پاسخ ماژولار رقابتی پرداخته شده است.

روش نخست استفاده از چندل مدل سیستم­های پرسش و پاسخ است که به صورت مستقل در یک فرآیند رقابتی برای کسب تخصص آموزش می­ببینند. روش دوم، استفاده از چندل مدل سیستم­های پرسش و پاسخ است که علاوه بر تبادل اطلاعات با یکدیگر در یک فرآیند رقابتی برای کسب تخصص آموزش می­بینند. در نهایت در روش سوم، متخصصین در یک ساختار به نام رمزگذار خودکار متغیر چیده می­شوند و بر خلاف دو روش پیشین به تعداد متخصصین، مدل وجود ندارد و تنها یک مدل رمزگذار خودکار متغیر وجود دارد که متخصصین در آن چیده می­شوند.

**منابع**

[1] O. Kolomiyets and M.-F. Moens, "A survey on question answering technology from an information retrieval perspective," Information Sciences, vol. 181, pp. 5412-5434, 2011.

[2] L. E. S. GREEN, E. C. BERKELEY, and C. GOTLIEB, "Conversation with a computer," Computers and Automation, vol. 8, 1959.

[3] B. F. Green;, A. K. Wolf;, C. Chomsky;, and K. Laughery;, "Baseball: an automatic question answerer, Computers & thought," MIT Press, Cambridge, 1963.

[4] W. W, K. R;, and W. B;, "The Lunar Sciences Natural Language Information System," BBN Report 2378, 1972.

[5] G. B.;, W. A., C. C;, and L. K.;, "Baseball: An Automatic Question Answer," 1983.

[6] T. Winograd, "Understanding natural language," Cognitive psychology, vol. 3, pp. 1-191, 1972.

[7] C. K. Riesbeck, R. C. Schank, N. M. Goldman, and C. J. Rieger III, "Inference and paraphrase by computer," Journal of the ACM (JACM), vol. 22, pp. 309-328, 1975.

[8] E. H. Shortliffe, R. Davis, S. G. Axline, B. G. Buchanan, C. C. Green, and S. N. Cohen, "Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system," Computers and biomedical research, vol. 8, pp. 303-320, 1975.

[9] V. Tablan, D. Damljanovic, and K. Bontcheva, A natural language query interface to structured information: Springer, 2008.

[10] J. Kupiec, “Murax: A robust linguistic approach for question answering using an on-line encyclopedia,” in Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Association for Computing Machinery, 1993, p. 181–190.

[11] E. M. Voorhees, “The trec-8 question answering track report,” NIST, Tech. Rep., 1999.

[12] E. M. Voorhees, “Overview of the trec 2001 question answering track,” in In Proceedings of TREC-10, 2001, pp. 42–51.

[13] ——, “Overview of the TREC 2002 question answering track,” in Proceedings of The Eleventh Text REtrieval Conference, TREC 2002, Gaithersburg, Maryland, USA, November 19-22, 2002, ser. NIST Special Publication, vol. 500-251. National Institute of Standards and Technology (NIST), 2002.

[14] E. Voorhees, “Overview of the trec 2003 question answering track,” NIST, Tech. Rep., 2003.

[15] J. Kupiec, “Murax: A robust linguistic approach for question answering using an on-line encyclopedia,” in Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Association for Computing Machinery, 1993, p. 181–190.

[16] C. Kwok, O. Etzioni, O. Etzioni, and D. S. Weld, “Scaling question answering to the web,” ACM Transactions on Information Systems, vol. 19, no. 3, pp. 242–262, 2001.

[17] E. Brill, S. Dumais, and M. Banko, “An analysis of the AskMSR question-answering system,” in Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002). Association for Computational Linguistics, 2002, pp. 257– 264.

[18] Z. Zheng, “Answerbus question answering system,” in Proceedings of the Second International Conference on Human Language Technology Research, ser. HLT ’02. Morgan Kaufmann Publishers Inc., 2002, p. 399–404.

[19] V. C. John Burger, Claire Cardie et al., “Issues, tasks and program structures to roadmap research in question & answering (q &a,” NIST, Tech. Rep., 2001.

[20] O. Kolomiyets and M.-F. Moens, “A survey on question answering technology from an information retrieval perspective,” Inf. Sci., vol. 181, no. 24, pp. 5412–5434, 2011.

[21] A. Allam and M. Haggag, “The question answering systems: A survey,” International Journal of Research and Reviews in Information Sciences, pp. 211–221, 2012.

[22] A. Mishra and S. K. Jain, “A survey on question answering systems with classification,” J. King Saud Univ. Comput. Inf. Sci., vol. 28, no. 3, p. 345–361, 2016.

[23] S. M. Harabagiu, S. J. Maiorano, and M. A. Paundefinedca, “Open-domain textual question answering techniques,” Nat. Lang. Eng., vol. 9, no. 3, p. 231–267, 2003.

[24] M. Pas¸ca, “Open-domain question answering from large text collections,” Computational Linguistics, vol. 29, no. 4, pp. 665–667, 2003.

[25] Z. Huang, S. Xu, M. Hu, X. Wang, J. Qiu, Y. Fu, Y. Zhao, Y. Peng, and C. Wang, “Recent trends in deep learning based open domain textual question answering systems,” IEEE Access, vol. 8, pp. 94 341–94 356, 2020.

[26] T. Lei, Z. Shi, D. Liu, L. Yang, and F. Zhu, “A novel cnn based method for question classification in intelligent question answering,” in Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence. Association for Computing Machinery, 2018.

[27] W. Xia, W. Zhu, B. Liao, M. Chen, L. Cai, and L. Huang, “Novel architecture for long short-term memory used in question classification,” Neuro computing, vol. 299, pp. 20–31, 2018.

[28] K. Nishida, I. Saito, A. Otsuka, H. Asano, and J. Tomita, “Retrieve-and-read: Multi-task learning of information retrieval and reading comprehension,” in Proceedings of the 27th ACM International Conference on Information and Knowledge Management, ser. CIKM ’18. Association for Computing Machinery, 2018, p. 647–656.

[29] V. Karpukhin, B. O˘guz, S. Min, L. Wu, S. Edunov, D. Chen, and W.-t. Yih, “Dense passage retrieval for open-domain question answering,” arXiv preprint arXiv:2004.04906, 2020.

[30] O. Khattab, C. Potts, and M. Zaharia, “Relevance-guided Supervision for OpenQA with ColBERT,” 2020. [Online]. Available: http://arxiv.org/abs/2007.00814.

[31] K. M. Hermann, T. Koˇcisk ´ y, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom, “Teaching machines to read and comprehend,” in Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1. MIT Press, 2015, pp. 1693–1701.

[32] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, “SQuAD: 100,000+ questions for machine comprehension of text,” in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2016, pp. 2383–2392.

[33] T. Nguyen, M. Rosenberg, X. Song, J. Gao, S. Tiwary, R. Majumder, and L. Deng, “MS MARCO: A human generated machine reading comprehension dataset,” 2016.

[34] G. Lai, Q. Xie, H. Liu, Y. Yang, and E. H. Hovy, “RACE: largescale reading comprehension dataset from examinations,” CoRR, vol. abs/1704.04683, 2017.

[35] P. Rajpurkar, R. Jia, and P. Liang, “Know what you don’t know: Unanswerable questions for SQuAD,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Association for Computational Linguistics, 2018, pp. 784–789.

[36] J. Li, M. Liu, M.-Y. Kan, Z. Zheng, Z. Wang, W. Lei, T. Liu, and B. Qin, “Molweni: A challenge multiparty dialogues-based machine reading comprehension dataset with discourse structure,” 2020.

[37] M. J. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi, “Bidirectional attention flow for machine comprehension,” in 5th International Conference on Learning Representations, ICLR 2017. OpenReview.net, 2017.

[38] W. Wang, N. Yang, F. Wei, B. Chang, and M. Zhou, “Gated self-matching networks for reading comprehension and question answering,” in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL. Association for Computational Linguistics, 2017, pp. 189–198.

[39] A. W. Yu, D. Dohan, M. Luong, R. Zhao, K. Chen, M. Norouzi, and Q. V. Le, “Qanet: Combining local convolution with global self-attention for reading comprehension,” in International Conference on Learning Representations, ICLR. OpenReview.net, 2018.

[40] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: pretraining of deep bidirectional transformers for language understanding,” CoRR, vol. abs/1810.04805, 2018.

[41] D. Chen, A. Fisch, J. Weston, and A. Bordes, “Reading Wikipedia to answer open-domain questions,” in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2017, pp. 1870–1879.

[42] S. Wang, M. Yu, X. Guo, Z.Wang, T. Klinger,W. Zhang, S. Chang, G. Tesauro, B. Zhou, and J. Jiang, “R3: Reinforced ranker-reader for open-domain question answering,” in AAAI, 2018.

[43] R. Das, S. Dhuliawala, M. Zaheer, and A. McCallum, “Multi-step retriever-reader interaction for scalable open-domain question answering,” in International Conference on Learning Representations, 2019.

[44] K. Guu, K. Lee, Z. Tung, P. Pasupat, and M.-W. Chang, “Realm: Retrieval-augmented language model pre-training,” CoRR, 2020.

[45] M. Ding, C. Zhou, Q. Chen, H. Yang, and J. Tang, “Cognitive graph for multi-hop reading comprehension at scale,” in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2019, pp. 2694–2703.

[46] Y. Nie, S. Wang, and M. Bansal, “Revealing the importance of semantic retrieval for machine reading at scale,” in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, 2019, pp. 2553–2566.

[47] Y. Lin, H. Ji, Z. Liu, and M. Sun, “Denoising distantly supervised open-domain question answering,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2018, pp. 1736–1745.

[48] J. Lee, S. Yun, H. Kim, M. Ko, and J. Kang, “Ranking paragraphs for improving answer recall in open-domain question answering,” in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2018, pp. 565–569.

[49] K. Lee, M.-W. Chang, and K. Toutanova, “Latent retrieval for weakly supervised open domain question answering,” in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2019, pp. 6086–6096.

[50] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. K¨ uttler, M. Lewis, W.-t. Yih, T. Rockt¨aschel, S. Riedel, and D. Kiela, “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” 2020. [Online]. Available: http://arxiv.org/abs/2005.11401

[51] C. yew Lin. Rouge, “A package for automatic evaluation of summaries,” pages 25–26, 2004.

[52] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. Bleu: A method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL ’02, pages 311–318, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.

[53] W. B. Croft, D. Metzler, and T. Strohman, Search engines: Information retrieval in practice: Addison-Wesley Reading, 2010.

[54] E. Brill, J. J. Lin, M. Banko, S. T. Dumais, and A. Y. Ng, "Data-Intensive Question Answering," in TREC, 2001.

[55] L. Dali, D. Rusu, B. Fortuna, D. Mladenic, and M. Grobelnik, "Question answering based on semantic graphs," in Proceedings of the workshop on semantic search (sem-search 2009), 2009.

[56] X. Li", Syntactic features in question answering," in Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, 2003, pp. 383-384.

[57] D. Moldovan, C. Clark, S. Harabagiu, and D. Hodges, "Cogex: A semantically and contextually enriched logic prover for question answering," Journal of Applied Logic, vol. 5, pp. 49-69, 2007.

[58] B. Katz, S. Felshin, D. Yuret, A. Ibrahim, J. Lin, G. Marton, et al., "Omnibase: Uniform access to heterogeneous data for question answering," in Natural Language Processing and Information Systems, ed: Springer, 2002, pp. 230-234.

[58] E. Lloret, H. Llorens, P. Moreda, E. Saquete, and M. Palomar, "Text summarization contribution to semantic question answering: New approaches for finding answers on the web," International Journal of Intelligent Systems, vol. 26, pp. 1125-1152, 2011.

[59] P. Spyns and J. Odijk, Essential Speech and Language Technology for Dutch: Springer, 2013.

[60] J.-M. Torres-Moreno, P.-L. St-Onge, M. Gagnon, M. El-Beze, and P. Bellot, "Automatic summarization system coupled with a question-answering system (qaas)," arXiv preprint arXiv:0905.2990, 2009.

[61] M. Biryukov, R. Angheluta, and M.-F. Moens, "Multidocument Question Answering Text Summarization Using Topic Signatures," JDIM, vol. 3, pp. 27-33, 2005.

[62] D. Demner-Fushman and J. Lin, "Answer extraction, semantic clustering, and extractive summarization for clinical question answering," in Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, 2006, pp. 841-848.

[63] Z. Shi, G. Melli, Y. Wang, Y. Liu, B. Gu, M. M. Kashani, et al., "Question answering summarization of multiple biomedical documents," in Advances in Artificial Intelligence, ed: Springer, 2007, pp. 284-295.

[64] D. Mollá and J. L. Vicedo, "Question answering in restricted domains: An overview," Computational Linguistics, vol. 33, pp. 41-61, 2007.

[65] R. Navigli, "Word sense disambiguation: A survey," ACM Computing Surveys (CSUR), vol. 41, p. 10, 2009.

[66] D. Demner-Fushman and J. Lin, "Answering clinical questions with knowledge-based and statistical techniques," Computational Linguistics, vol. 33, pp. 63-103, 2007.

[67] S. M. Harabagiu, S. J. Maiorano, and M. A. Paundefinedca, “Open-domain textual question answering techniques,” Nat. Lang. Eng., vol. 9, no. 3, p. 231–267, 2003.

[68] M. Pas¸ca, “Open-domain question answering from large text collections,” Computational Linguistics, vol. 29, no. 4, pp. 665–667, 2003.

[69] E. M. Voorhees, “The trec-8 question answering track report,” NIST, Tech. Rep., 1999.

[70] D. Moldovan, M. Pasca, S. Harabagiu, and M. Surdeanu, “Performance issues and error analysis in an open-domain question answering system,” in Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2002, pp. 33–40.

[71] O. Tsur, M. de Rijke, and K. Sima'an, "Biographer: Biography questions as a restricted domain question answering task," 2004.

[72] L. Ratinov and D. Roth, "Design challenges and misconceptions in named entity recognition," in Proceedings of the Thirteenth Conference on Computational Natural Language Learning, 2009, pp. 147-155.

[73] U. Hermjakob, "Parsing and question classification for question answering," in Proceedings of the workshop on Open-domain question answering-Volume 12, 2 ,444pp. 1-6.

[74] T. Strzalkowski and S. M. Harabagiu, Advances in open domain question answering vol. 32: Springer, 2006.

[75] D. Moldovan, M. Paşca, S. Harabagiu, and M. Surdeanu, "Performance issues and error analysis in an open-domain question answering system," ACM Transactions on Information Systems (TOIS), vol. 21, pp. 133-154, 2003.

[76] B. Loni, G. Van Tulder, P. Wiggers, D. M. Tax, and M. Loog, "Question classification by weighted combination of lexical, syntactic and semantic features," in Text, Speech and Dialogue, 2011, pp. 243-250.

[77] X. Li and D. Roth, "Learning question classifiers," in Proceedings of the 19th international conference on Computational linguistics-Volume 1, 2002, pp. 1-7.

[78] X. Li and D. Roth, “Learning question classifiers,” in COLING 2002: The 19th International Conference on Computational Linguistics, 2002.

[79] J. Suzuki, H. Taira, Y. Sasaki, and E. Maeda, “Question classification using HDAG kernel,” in Proceedings of the ACL 2003 Workshop on Multilingual Summarization and Question Answering. Association for Computational Linguistics, 2003, pp. 61–68.

[80] M. Collins, "Head-driven statistical models for natural language parsing," Computational linguistics, vol. 29, pp. 589-637, 2003.

[81] H. Lee, A. Chang, Y. Peirsman, N. Chambers, M. Surdeanu, and D. Jurafsky, "Deterministic coreference resolution based on entity-centric, precision-ranked rules," Computational Linguistics, vol. 39, pp. 885 .4444 ,446-

[82] T. Zhao, X. Lu, and K. Lee, “Sparta: Efficient open-domain question answering via sparse transformer matching retrieval,” arXiv preprint arXiv:2009.13013, 2020.

[83] D. Radev, W. Fan, H. Qi, H. Wu, and A. Grewal, "Probabilistic question answering on the web," Journal of the American Society for Information Science and Technology, vol. 56, pp. 571-583, 2005.

[84] B. Loni, G. Van Tulder, P. Wiggers, D. M. Tax, and M. Loog, "Question classification by weighted combination of lexical, syntactic and semantic features," in Text, Speech and Dialogue, 2011, pp. 243-250.

[85] H. Sundblad, "Question Classification in Question Answering Systems," 2007.

[86] X. Li and D. Roth", Learning question classifiers: the role of semantic information," Natural Language Engineering, vol. 12, pp. 229-249, 2006.

[87] S. K. Ray, S. Singh, and B. P. Joshi, "A semantic approach for question classification using WordNet and Wikipedia," Pattern Recognition Letters, vol. 31, pp. 1935-1943, 2010.

[88] J. Silva, L. Coheur, A. C. Mendes, and A. Wichert, "From symbolic to sub-symbolic information in question classification," Artificial Intelligence Review, vol. 35, pp. 137-154, 2011.

[89] Z. Qi, Y. Tian, and Y. Shi, "Robust twin support vector machine for pattern classification," Pattern Recognition, vol. 46, pp. 305-316, 2013.

[90] M. Vargas-Vera and E. Motta, "AQUA: A question answering system for heterogeneous sources," Knowledge Media Institute the Open University [kmi. open. ac. uk/publications/pdf/kmi-04-20. pdf], 2004.

[91] O. Tsur, M. de Rijke, and K. Sima'an, "Biographer: Biography questions as a restricted domain question answering task," 2004.

[92] F. Benamara, "Cooperative question answering in restricted domains: the WEBCOOP experiment," in Proceedings of the Workshop Question Answering in Restricted Domains, within ACL, 2004.

[93] Y. Niu and G. Hirst, "Analysis of semantic classes in medical text for question answering," in Proceedings of the ACL 2004 Workshop on Question Answering in Restricted Domains, 2004, pp. 54-61.

[94] D. Demner-Fushman and J. Lin, "Knowledge extraction for clinical question answering: Preliminary results," in Proceedings of the AAAI-05 Workshop on Question Answering in Restricted Domains, 2005, pp. 9-13.

[95] H. Yu, C. Sable, and H. R. Zhu, "Classifying medical questions based on an evidence taxonomy," in Proceedings of the AAAI 2005 workshop on question answering in restricted domains, 2005.

[96] B. W. van Schooten, R. op den Akker, S. Rosset, O. Galibert, A. Max, and G. Illouz, "Follow-up question handling in the IMIX and Ritel systems: A comparative study," Natural Language Engineering, vol. 15, pp. 97-118, 2009.

[97] D. Ferrés and H. Rodríguez, "Experiments adapting an open-domain question answering system to the geographical domain using scope-based resources," in Proceedings of the Workshop on Multilingual Question Answering, 2006, pp. 69-76.

[98] A. Defazio, "Natural Language Question Answering Over Triple Knowledge Bases," 2009.

[99] D. Vickrey and D. Koller, "Sentence Simplification for Semantic Role Labeling," in ACL, 2008, pp .454-444.

[100] M.-C. De Marneffe, B. MacCartney, and C. D. Manning, "Generating typed dependency parses from phrase structure parses," in Proceedings of LREC, 2006, pp. 449-454.

[101] V. Punyakanok, D. Roth, and W.-t. Yih, "The importance of syntactic parsing and inference in semantic role labeling," Computational Linguistics, vol. 34, pp. 257-287, 2008.

[102] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1, 2003, pp. 173-180.

[103] Y. Pan, Y. Tang, L. Lin, and Y. Luo, "Question classification with semantic tree kernel," in Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, 2008, pp. 837-838.

[104] F. Rinaldi, J. Dowdall, G. Schneider, and A. Persidis, "Answering questions in the genomics domain," in Proceedings of the ACL 2004 Workshop on Question Answering in Restricted Domains, 2004, pp. 46-53.

[105] V. Krishnan, S. Das, and S. Chakrabarti, "Enhanced answer type inference from questions using sequential models," in Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, 2005, pp. 315-322.

[106] C. Wilson, G. Candela, and C. Watson, "Neural network fingerprint classification," Journal of Artificial Neural Networks, vol. 1, pp. 203-228, 1994.

[107] E. J. Pettit and M. J. Pettit, "Analysis of the performance of a genetic algorithm-based system for message classification in noisy environments," International Journal of Man-Machine Studies, vol. 27, pp. 205-220, 1987.

[108] M. A. Friedl and C. E. Brodley, "Decision tree classification of land cover from remotely sensed data," Remote sensing of environment, vol. 61, pp. 399-409, 1997.

[109] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler", Support vector machine classification and validation of cancer tissue samples using microarray expression data," Bioinformatics, vol. 16, pp. 906-914, 2000.

[110] T. Lei, Z. Shi, D. Liu, L. Yang, and F. Zhu, “A novel cnnbased method for question classification in intelligent question answering,” in Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence. Association for Computing Machinery, 2018.

[111] W. Xia, W. Zhu, B. Liao, M. Chen, L. Cai, and L. Huang, “Novel architecture for long short-term memory used in question classification,” Neurocomputing, vol. 299, pp. 20–31, 2018.

[112] V. Karpukhin, B. O˘guz, S. Min, L. Wu, S. Edunov, D. Chen, and W.-t. Yih, “Dense passage retrieval for open-domain question answering,” arXiv preprint arXiv:2004.04906, 2020.

[113] R. Das, S. Dhuliawala, M. Zaheer, and A. McCallum, “Multi-step retriever-reader interaction for scalable open-domain question answering,” in International Conference on Learning Representations, 2019.

[114] K. Guu, K. Lee, Z. Tung, P. Pasupat, and M.-W. Chang, “Realm: Retrieval-augmented language model pre-training,” CoRR, 2020.

[115] Y. Feldman et al., “Multi-hop paragraph retrieval for opendomain question answering,” in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2019, pp. 2296–2309.

[116] P. Ram and A. G. Gray, “Maximum inner-product search using cone trees,” in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012, pp. 931–939.

[117] A. Shrivastava and P. Li, “Asymmetric lsh (alsh) for sublinear time maximum inner product search (mips),” in Advances in Neural Information Processing Systems, 2014, pp. 2321–2329.

[118] F. Shen, W. Liu, S. Zhang, Y. Yang, and H. Tao Shen, “Learning binary codes for maximum inner product search,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 4148–4156.

[119] K. M. Hermann, T. Koˇcisk ´ y, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom, “Teaching machines to read and comprehend,” in Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1. MIT Press, 2015, pp. 1693–1701.

[120] T. Nguyen, M. Rosenberg, X. Song, J. Gao, S. Tiwary, R. Majumder, and L. Deng, “MS MARCO: A human generated machine reading comprehension dataset,” 2016.

[121] G. Lai, Q. Xie, H. Liu, Y. Yang, and E. H. Hovy, “RACE: largescale reading comprehension dataset from examinations,” CoRR, vol. abs/1704.04683, 2017.

[122] P. Rajpurkar, R. Jia, and P. Liang, “Know what you don’t know: Unanswerable questions for SQuAD,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Association for Computational Linguistics, 2018, pp. 784–789.

[123] Yin Zhang, Rong Jin, Zhi-Hua Zhou, “Understanding bag-of-words model: A statistical framework,” in International Journal of Machine Learning and Cybernetics 2010, pp. 43-52.

[124] W. Wang, N. Yang, F. Wei, B. Chang, and M. Zhou, “Gated self-matching networks for reading comprehension and question answering,” in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL. Association for Computational Linguistics, 2017, pp. 189–198.

[125] A. W. Yu, D. Dohan, M. Luong, R. Zhao, K. Chen, M. Norouzi, and Q. V. Le, “Qanet: Combining local convolution with global self-attention for reading comprehension,” in International Conference on Learning Representations, ICLR. OpenReview.net, 2018.

[126] Kim, SW., Gil, JM. Research paper classification systems based on TF-IDF and LDA schemes. Hum. Cent. Comput. Inf. Sci. 9, 30 (2019). https://doi.org/10.1186/s13673-019-0192-7.

[127] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA, 2017, pp. 5998–6008.

[128] M. Dehghani, H. Azarbonyad, J. Kamps, and M. de Rijke, “Learning to transform, combine, and reason in open-domain question answering,” in Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, ser. WSDM ’19. Association for Computing Machinery, 2019, p. 681–689.

[129] B. Dhingra, M. Zaheer, V. Balachandran, G. Neubig, R. Salakhutdinov, and W. W. Cohen, “Differentiable reasoning over a virtual knowledge base,” in International Conference on Learning Representations, 2020.

[130] B. Kratzwald, A. Eigenmann, and S. Feuerriegel, “RankQA: Neural question answering with answer re-ranking,” in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2019, pp. 6076–6085.

[131] B. Kratzwald et al., “Adaptive document retrieval for deep question answering,” in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2018, pp. 576–581.

[132] W. Yang, Y. Xie, A. Lin, X. Li, L. Tan, K. Xiong, M. Li, and J. Lin, “End-to-end open-domain question answering with BERTserini,” in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations). Association for Computational Linguistics, 2019, pp. 72–77.

[133] Z. Wang, P. Ng, X. Ma, R. Nallapati, and B. Xiang, “Multi-passage BERT: A globally normalized BERT model for open-domain question answering,” in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, 2019, pp. 5878–5882.

[134] K. Weinberger, A. Dasgupta, J. Langford, A. Smola, and J. Attenberg, “Feature hashing for large scale multitask learning,” in Proceedings of the 26th Annual International Conference on Machine Learning. Association for Computing Machinery, 2009, p. 1113–1120.

[135] P. Yang, H. Fang, and J. Lin, “Anserini: Enabling the use of lucene for information retrieval research,” in Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, ser. SIGIR ’17. Association for Computing Machinery, 2017, p. 1253–1256.

[136] M. Seo, J. Lee, T. Kwiatkowski, A. Parikh, A. Farhadi, and H. Hajishirzi, “Real-time open-domain question answering with densesparse phrase index,” in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2019, pp. 4430–4441.

[137] [Stephen E. Robertson](https://www.researchgate.net/profile/Stephen-Robertson-11?_sg%5B0%5D=rbOQrdOtYUk0sgeP7DGj89-5_tBpY8vJ2I-6Epeob3p64b0zZ20IqiQHNXjEoaMohC5XNEk.AVjoMWebPgUiJCF2i6nVhXKkMYPiAn6CIUEoERADbw3aUCMtNRqs519Z0OYfcsfhW21rm9ctlQ0uPorqCwhf3g&_sg%5B1%5D=2_wvEWuQUPJZDAhHlSBuVxM9lnscGlb2pPnCRyB8k3f3WhaLy3hV0xb9z1dtyc9xkI8fLIM.QgGrX2grXT3WtjHxgJ2qHH6YSRBRz176SS8gJ4kMKx7q3B84FAUzwTJgV0GPEFwaTcHth4glJnAfLSUEMpTQGQ), [Hugo Zaragoza](https://www.researchgate.net/profile/Hugo-Zaragoza?_sg%5B0%5D=rbOQrdOtYUk0sgeP7DGj89-5_tBpY8vJ2I-6Epeob3p64b0zZ20IqiQHNXjEoaMohC5XNEk.AVjoMWebPgUiJCF2i6nVhXKkMYPiAn6CIUEoERADbw3aUCMtNRqs519Z0OYfcsfhW21rm9ctlQ0uPorqCwhf3g&_sg%5B1%5D=2_wvEWuQUPJZDAhHlSBuVxM9lnscGlb2pPnCRyB8k3f3WhaLy3hV0xb9z1dtyc9xkI8fLIM.QgGrX2grXT3WtjHxgJ2qHH6YSRBRz176SS8gJ4kMKx7q3B84FAUzwTJgV0GPEFwaTcHth4glJnAfLSUEMpTQGQ), “The Probabilistic Relevance Framework: BM25 and Beyond,” in Foundations and Trends® in Information Retrieval, 2009, pp. 333-389.

[138] Y. Zhang, P. Nie, X. Geng, A. Ramamurthy, L. Song, and D. Jiang, “Dc-bert: Decoupling question and document for efficient contextual encoding,” 2020.

[139] O. Khattab and M. Zaharia, “Colbert: Efficient and effective passage search via contextualized late interaction over bert,” in Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, ser. SIGIR ’20. Association for Computing Machinery, 2020, p. 39–48.

[140] P. Qi, X. Lin, L. Mehr, Z. Wang, and C. D. Manning, “Answering complex open-domain questions through iterative query generation,” in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, 2019, pp. 2590–2602.

[141] W. Xiong, X. L. Li, S. Iyer, J. Du, P. Lewis, W. Y. Wang, Y. Mehdad, W.-t. Yih, S. Riedel, D. Kiela et al., “Answering complex opendomain questions with multi-hop dense retrieval,” arXiv preprint arXiv:2009.12756, 2020.

[142] Y. Zhang, P. Nie, A. Ramamurthy, and L. Song, “Ddrqa: Dynamic document reranking for open-domain multi-hop question answering,” arXiv preprint arXiv:2009.07465, 2020.

[143] Y. Mao, P. He, X. Liu, Y. Shen, J. Gao, J. Han, and W. Chen, “Generation-Augmented Retrieval for Open-domain Question Answering,” 2020. [Online]. Available: http://arxiv.org/abs/2009.08553

[144] A. Asai, K. Hashimoto, H. Hajishirzi, R. Socher, and C. Xiong, “Learning to retrieve reasoning paths over wikipedia graph for question answering,” in International Conference on Learning Representations, 2020.

[145] S. Min, D. Chen, L. Zettlemoyer, and H. Hajishirzi, “Knowledge Guided Text Retrieval and Reading for Open Domain Question Answering,” 2019. [Online]. Available: http://arxiv.org/abs/1911.03868

[146] Z. Yang, P. Qi, S. Zhang, Y. Bengio, W. Cohen, R. Salakhutdinov, and C. D. Manning, “HotpotQA: A dataset for diverse, explainable multi-hop question answering,” in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2018, pp. 2369–2380.

[147] J. Welbl, P. Stenetorp, and S. Riedel, “Constructing datasets for multi-hop reading comprehension across documents,” Transactions of the Association for Computational Linguistics, pp. 287–302, 2018. [Online]. Available: https://www.aclweb.org/anthology/Q18-1021

[148] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, “BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension,” in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020, pp. 7871–7880.

[149] K. Cho, B. van Merrienboer, C¸ . G¨ ulc¸ehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” in EMNLP. ACL, 2014, pp. 1724–1734.

[150] A. Lampert, “A quick introduction to question answering,” Dated December, 2004.

[151] P. M. Htut, S. Bowman, and K. Cho, “Training a ranking function for open-domain question answering,” in Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop. Association for Computational Linguistics, 2018, pp. 120–127.

[152] P. Banerjee, K. K. Pal, A. Mitra, and C. Baral, “Careful selection of knowledge to solve open book question answering,” in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2019, pp. 6120–6129.

[153] J. Wang, A. Jatowt, M. F¨arber, and M. Yoshikawa, “Answering event-related questions over long-term news article archives,” in ECIR, ser. Lecture Notes in Computer Science, vol. 12035. Springer, 2020, pp. 774–789.

[154] J. Wang, A. Jatowt, M. F¨arber, and M. Yoshikawa, “Improving question answering for event-focused questions in temporal collections of news articles,” Information Retrieval Journal, vol. 24, no. 1, pp. 29–54, 2021.

[155] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, and A. Bordes, “Supervised learning of universal sentence representations from natural language inference data,” in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2017, pp. 670–680.

[156] A. Santoro, D. Raposo, D. G. Barrett, M. Malinowski, R. Pascanu, P. Battaglia, and T. Lillicrap, “A simple neural network module for relational reasoning,” in Advances in Neural Information Processing Systems 30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4967–4976.

[157] M. Richardson, C. J. Burges, and E. Renshaw, “MCTest: A challenge dataset for the open-domain machine comprehension of text,” in Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2013, pp. 193–203.

[158] D. Dua, Y. Wang, P. Dasigi, G. Stanovsky, S. Singh, and M. Gardner, “DROP: A reading comprehension benchmark requiring discrete reasoning over paragraphs,” in Proc. of NAACL, 2019.

[159] G. Izacard and E. Grave, “Leveraging passage retrieval with generative models for open domain question answering,” arXiv preprint arXiv:2007.01282, 2020.

[160] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in ICLR, 2017.

[161] C. Clark and M. Gardner, “Simple and effective multi-paragraph reading comprehension,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL. Association for Computational Linguistics, 2018, pp. 845–855.

[162] C. Tan, F. Wei, N. Yang, B. Du, W. Lv, and M. Zhou, “S-net: From answer extraction to answer synthesis for machine reading comprehension,” in AAAI. AAAI Press, 2018, pp. 5940–5947.

[163] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, “Exploring the limits of transfer learning with a unified text-to-text transformer,” arXiv e-prints, 2019.

[164] S. Liu, X. Zhang, S. Zhang, H. Wang, and W. Zhang, “Neural machine reading comprehension: Methods and trends,” CoRR, vol. abs/1907.01118, 2019.

[165] S. Wang, M. Yu, J. Jiang, W. Zhang, X. Guo, S. Chang, Z. Wang, T. Klinger, G. Tesauro, and M. Campbell, “Evidence aggregation for answer re-ranking in open-domain question answering,” in 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings. ICLR, 2018.

[166] S. Wang and J. Jiang, “Learning natural language inference with LSTM,” in Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. The Association for Computational Linguistics, 2016, pp. 1442–1451.

[167] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, “Language models are unsupervised multitask learners,” OpenAI blog, vol. 1, no. 8, p. 9, 2019.

[168] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell et al., “Language models are few-shot learners,” arXiv preprint arXiv:2005.14165, 2020.

[169] J. Johnson, M. Douze, and H. J´egou, “Billion-scale similarity search with gpus,” CoRR, vol. abs/1702.08734, 2017.

[170] F. Petroni, T. Rockt¨aschel, P. Lewis, A. Bakhtin, Y. Wu, A. H. Miller, and S. Riedel, “Language models as knowledge bases?” arXiv preprint arXiv:1909.01066, 2019.

[171] A. Roberts, C. Raffel, and N. Shazeer, “How much knowledge can you pack into the parameters of a language model?” arXiv preprint arXiv:2002.08910, 2020.

[172] M. Seo, T. Kwiatkowski, A. Parikh, A. Farhadi, and H. Hajishirzi, “Phrase-indexed question answering: A new challenge for scalable document comprehension,” in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2018, pp. 559–564.

[173] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, “Dbpedia: A nucleus for a web of open data,” in The Semantic Web. Springer Berlin Heidelberg, 2007, pp. 722–735.

[174] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, “Freebase: A collaboratively created graph database for structuring human knowledge,” in Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. ACM, 2008, pp. 1247–1250.

[175] J. Hoffart, F. M. Suchanek, K. Berberich, and G. Weikum, “Yago2: A spatially and temporally enhanced knowledge base from wikipedia,” Artif. Intell., vol. 194, pp. 28–61, 2013.

[176] H. Sun, H. Ma, W.-t. Yih, C.-T. Tsai, J. Liu, and M.-W. Chang, “Open domain question answering via semantic enrichment,” in Proceedings of the 24th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2015, pp. 1045–1055.

[177] H. Sun, B. Dhingra, M. Zaheer, K. Mazaitis, R. Salakhutdinov, and W. Cohen, “Open domain question answering using early fusion of knowledge bases and text,” in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2018, pp. 4231–4242.

[178] H. Sun, T. Bedrax-Weiss, and W. Cohen, “PullNet: Open domain question answering with iterative retrieval on knowledge bases and text,” in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, 2019, pp. 2380–2390.

[179] Y. Li, D. Tarlow, M. Brockschmidt, and R. S. Zemel, “Gated graph sequence neural networks,” in ICLR, 2016.

[180] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, “The graph neural network model,” IEEE Transactions on Neural Networks, vol. 20, no. 1, pp. 61–80, 2009.

[181] D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, N. Schlaefer, and C. Welty, “Building watson: An overview of the deepqa project,” AI Magazine, vol. 31, no. 3, pp. 59–79, 2010.

[182] J. Gao, M. Galley, and L. Li, “Neural approaches to conversational ai,” Foundations and Trends® in Information Retrieval, vol. 13, no. 2-3, pp. 127–298, 2019.

[183] W. Lei, X. He, M. de Rijke, and T.-S. Chua, “Conversational recommendation: Formulation, methods, and evaluation,” in Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, ser. SIGIR ’20. Association for Computing Machinery, 2020, p. 2425–2428.

[184] H. Zhu, L. Dong, F. Wei, W. Wang, B. Qin, and T. Liu, “Learning to ask unanswerable questions for machine reading comprehension,” in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2019, pp. 4238–4248.

[185] M. Hu, F. Wei, Y. xing Peng, Z. X. Huang, N. Yang, and M. Zhou, “Read + verify: Machine reading comprehension with unanswerable questions,” ArXiv, vol. abs/1808.05759, 2018.

[186] M. Aliannejadi, H. Zamani, F. Crestani, and W. B. Croft, “Asking clarifying questions in open-domain information-seeking conversations,” in Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Association for Computing Machinery, 2019, p. 475–484.

[187] X. Du, J. Shao, and C. Cardie, “Learning to ask: Neural question generation for reading comprehension,” in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver, Canada: Association for Computational Linguistics, 2017, pp. 1342–1352.

[188] N. Duan, D. Tang, P. Chen, and M. Zhou, “Question generation for question answering,” in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark: Association for Computational Linguistics, 2017, pp. 866–874. [Online]. Available: <https://www.aclweb.org/anthology/D17-1090>.

[189] Q. Zhou, N. Yang, F. Wei, C. Tan, H. Bao, and M. Zhou, “Neural question generation from text: A preliminary study,” CoRR, vol. abs/1704.01792, 2017.

[190] E. Choi, H. He, M. Iyyer, M. Yatskar, W. Yih, Y. Choi, P. Liang, and L. Zettlemoyer, “Quac: Question answering in context,” CoRR, vol. abs/1808.07036, 2018.

[191] C. C. Chen Qu, Liu Yang et al., “Open-retrieval conversational question answering,” CoRR, vol. abs/2005.11364, 2020.

[192] S. Reddy, D. Chen, and C. D. Manning, “Coqa: A conversational question answering challenge,” CoRR, vol. abs/1808.07042, 2018.

[193] Fengbin Zhu, Wenqiang Lei\*, Chao Wang, Jianming Zheng, Soujanya Poria, Tat-Seng Chua, “Retrieving and Reading: A Comprehensive Survey on Open-domain Question Answering,” [arXiv:2101.00774v3](https://arxiv.org/abs/2101.00774v3), 2021.

[194] Xilinx. ["HDL Synthesis for FPGAs Design Guide"](http://www.xilinx.com/txpatches/pub/documentation/xactstep6/hdlsynth.pdf). section 3.13: "Encoding State Machines". Appendix A: "Accelerate FPGA Macros with One-Hot Approach". 1995.

[195] [Tomas Mikolov](https://arxiv.org/search/cs?searchtype=author&query=Mikolov%2C+T), [Kai Chen](https://arxiv.org/search/cs?searchtype=author&query=Chen%2C+K), [Greg Corrado](https://arxiv.org/search/cs?searchtype=author&query=Corrado%2C+G), [Jeffrey Dean](https://arxiv.org/search/cs?searchtype=author&query=Dean%2C+J), “Efficient Estimation of Word Representations in Vector Space,” arXiv preprint arXiv:1301.3781, 2013 - arxiv.org.

[196] [Jeffrey Pennington](https://paperswithcode.com/author/jeffrey-pennington), [Richard Socher](https://paperswithcode.com/author/richard-socher), [Christopher Manning](https://paperswithcode.com/author/christopher-manning), “GloVe: Global Vectors for Word Representation,” in [Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)](https://aclanthology.org/volumes/D14-1/), 2014, pp. 1532-1543.

[197] Dirk P. Kroese, Reuven Y. Rubinstein, Izack Cohen, Sergey Porotsk, Thomas Taimre, “The Cross Entropy method for Fast Policy Search,” in Encyclopedia of Operations Research and Management Sciences, 2012. Edition: Third edition Chapter: Cross-Entropy Method Publisher: Springer-Verlag.

[198] Jong-Hoon Won1, Thomas Pany2, Bernd Eissfeller, “A Maximum Likelihood Estimator Based Tracking Algorithm for GNSS Signals,” [Proceedings of the Korean Institute of Navigation and Port Research Conference,](https://www.koreascience.or.kr/journal/GHMHAH.1ff8page) Volume 2, 2006, pp. 15-22.

[199] N. Wiener, "The theory of prediction, " Modern mathematics for engineers, vol. 1, pp. 165-183, 1956.

[200] C. W. J. Granger, “Investigating Causal Relations by Econometric Models and Cross-spectral Methods,” Published By: The Econometric Society, [Vol. 37, No. 3 (Aug., 1969)](https://www.jstor.org/stable/i332739), pp. 424-438.

[201] Ishida, R., Torisawa, K., Oh, J. H., Iida, R., Kruengkrai, C., Kloetzer, J.: Semidistantly supervised neural model for generating compact answers to open-domain why questions. In Thirty-Second AAAI Conference on Artificial Intelligence (2018).

[202] Fukumoto, J. I.: Question Answering System for Non-Factoid Type Questions and Automatic Evaluation based on BE Method. In NTCIR, 2007.

[203] Oh, J. H., Torisawa, K., Hashimoto, C., Sano, M., De Saeger, S., & Ohtake, K.: Why-question answering using intra-and inter-sentential causal relations. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Volume 1: Long Papers, 2013, pp. 1733-1743.

[204] Higashinaka, R., & Isozaki, H.: Corpus-based question answering for why- questions. In Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I, 2008.

[205] Johnson, C. R., Schwarzer-Petruck, M., Baker, C. F., Ellsworth, M., Ruppenhofer, J., & Fill- more, C. J. Framenet: Theory and practice, 2003.

[206] Palmer, M.: Semlink: Linking propbank, verbnet and framenet. In Proceedings of the generative lexicon conference, Pisa, Italy: GenLex-09 ,2009, pp. 9-15.

[207] Yokoi, T.: The EDR electronic dictionary. Communications of the ACM, 38(11), 1995, pp. 42- 44.

1. - Bag of words [↑](#footnote-ref-1)
2. - Knowledge Base [↑](#footnote-ref-2)
3. - (Closed/Restricted-Domain) Question Answering Systems [↑](#footnote-ref-3)
4. - Open-Domain Question Answering Systems [↑](#footnote-ref-4)
5. - Long Short-term memory [↑](#footnote-ref-5)
6. - Gated Recurrent Unit [↑](#footnote-ref-6)
7. - Autoencoders [↑](#footnote-ref-7)
8. - Google [↑](#footnote-ref-8)
9. - Bing [↑](#footnote-ref-9)
10. - Machine Reading Comprehension (MRC) [↑](#footnote-ref-10)
11. - Imitation Game [↑](#footnote-ref-11)
12. - Turing Test [↑](#footnote-ref-12)
13. - Conversation Machin [↑](#footnote-ref-13)
14. - Green [↑](#footnote-ref-14)
15. - BASEBALL [↑](#footnote-ref-15)
16. - LUNAR [↑](#footnote-ref-16)
17. - Shallow Parsing [↑](#footnote-ref-17)
18. - Woods [↑](#footnote-ref-18)
19. - Wingorad [↑](#footnote-ref-19)
20. - SHRDLU [↑](#footnote-ref-20)
21. - MARGIE [↑](#footnote-ref-21)
22. - Schank [↑](#footnote-ref-22)
23. - MYCIN [↑](#footnote-ref-23)
24. - MURAX [↑](#footnote-ref-24)
25. - MUDLER [↑](#footnote-ref-25)
26. - Factoid [↑](#footnote-ref-26)
27. - Pipeline [↑](#footnote-ref-27)
28. - Lei [↑](#footnote-ref-28)
29. - Xia [↑](#footnote-ref-29)
30. - Covolutional Neural Network [↑](#footnote-ref-30)
31. - Long-Short Term Memory [↑](#footnote-ref-31)
32. - Latent Space [↑](#footnote-ref-32)
33. - Retriever-Reader [↑](#footnote-ref-33)
34. - End to End [↑](#footnote-ref-34)
35. - Vectorization [↑](#footnote-ref-35)
36. - Sparse Vector Representations [↑](#footnote-ref-36)
37. - Dense Vector Representations [↑](#footnote-ref-37)
38. - Bag of Words (BoW) [↑](#footnote-ref-38)
39. - Gram [↑](#footnote-ref-39)
40. - Term Frequancy [↑](#footnote-ref-40)
41. - Inverse Document Frequancy [↑](#footnote-ref-41)
42. - Robertson [↑](#footnote-ref-42)
43. - Okapi [↑](#footnote-ref-43)
44. - تعداد واژه­ها [↑](#footnote-ref-44)
45. - احتمال رخداد یک واژه [↑](#footnote-ref-45)
46. - Word Embedding [↑](#footnote-ref-46)
47. - Global Vectors for Word Representation [↑](#footnote-ref-47)
48. - co-occurrence matrix [↑](#footnote-ref-48)
49. - BERT [↑](#footnote-ref-49)
50. - feature extraction [↑](#footnote-ref-50)
51. - Fine-Tuning [↑](#footnote-ref-51)
52. - Encoder [↑](#footnote-ref-52)
53. - BERT’s vocabulary [↑](#footnote-ref-53)
54. - exact match [↑](#footnote-ref-54)
55. - Recall-Oriented Understudy for Gisting Evaluation [↑](#footnote-ref-55)
56. - Bilingual Evaluation Understudy [↑](#footnote-ref-56)
57. - Cloze test [↑](#footnote-ref-57)
58. - Confusion Matrix [↑](#footnote-ref-58)
59. - Loss Function [↑](#footnote-ref-59)
60. - Claude Shannon [↑](#footnote-ref-60)
61. - Maximum Likelihood Estimator [↑](#footnote-ref-61)
62. - Bag of words [↑](#footnote-ref-62)
63. - W. B. Croft [↑](#footnote-ref-63)
64. - Stop Words [↑](#footnote-ref-64)
65. - Morphology [↑](#footnote-ref-65)
66. - Kolomiyets [↑](#footnote-ref-66)
67. - Part-of-speech(POS) [↑](#footnote-ref-67)
68. - Chunker [↑](#footnote-ref-68)
69. - Parser [↑](#footnote-ref-69)
70. - Moldovan [↑](#footnote-ref-70)
71. - Machin Representation Language [↑](#footnote-ref-71)
72. - Dali [↑](#footnote-ref-72)
73. - Katz [↑](#footnote-ref-73)
74. - Start [↑](#footnote-ref-74)
75. - Biryukov [↑](#footnote-ref-75)
76. - Demner-Fushman [↑](#footnote-ref-76)
77. - Shi [↑](#footnote-ref-77)
78. - Concept Similarity [↑](#footnote-ref-78)
79. - Knowledge Base [↑](#footnote-ref-79)
80. - (Closed/Restricted-Domain) Question Answering Systems [↑](#footnote-ref-80)
81. - Open-Domain Question Answering Systems [↑](#footnote-ref-81)
82. - Brill [↑](#footnote-ref-82)
83. - Redundancy based [↑](#footnote-ref-83)
84. - Corpus [↑](#footnote-ref-84)
85. - Data\_Intensive [↑](#footnote-ref-85)
86. - Resolution [↑](#footnote-ref-86)
87. - Domain Context [↑](#footnote-ref-87)
88. - Navigli [↑](#footnote-ref-88)
89. - Word-Sense Disambiguation [↑](#footnote-ref-89)
90. - Additional Resource [↑](#footnote-ref-90)
91. - Strzalkowski [↑](#footnote-ref-91)
92. - Classification-Based Approaches [↑](#footnote-ref-92)
93. - Non-Classification-Based Approaches [↑](#footnote-ref-93)
94. - Location [↑](#footnote-ref-94)
95. - Li [↑](#footnote-ref-95)
96. - Roth [↑](#footnote-ref-96)
97. - Fine [↑](#footnote-ref-97)
98. - Coarse [↑](#footnote-ref-98)
99. - Zhao [↑](#footnote-ref-99)
100. - Collins [↑](#footnote-ref-100)
101. - SVM [↑](#footnote-ref-101)
102. - KNN [↑](#footnote-ref-102)
103. - Decision Tree [↑](#footnote-ref-103)
104. - Definition [↑](#footnote-ref-104)
105. - Naïve Bayes [↑](#footnote-ref-105)
106. - Decision Trees [↑](#footnote-ref-106)
107. - Sparse Network of Winnows [↑](#footnote-ref-107)
108. - Query Formulation [↑](#footnote-ref-108)
109. - POS [↑](#footnote-ref-109)
110. - mismatch [↑](#footnote-ref-110)
111. - Boolean Model [↑](#footnote-ref-111)
112. - Vector Space Models [↑](#footnote-ref-112)
113. - Probabilistic Models [↑](#footnote-ref-113)
114. - Language Models [↑](#footnote-ref-114)
115. - Punyakanok [↑](#footnote-ref-115)
116. - Okapi [↑](#footnote-ref-116)
117. - Query Likelihood Model [↑](#footnote-ref-117)
118. - De Marneffe [↑](#footnote-ref-118)
119. - Document Filtering [↑](#footnote-ref-119)
120. - Document Re-ranking [↑](#footnote-ref-120)
121. - Document Selection [↑](#footnote-ref-121)
122. - Factoid Question [↑](#footnote-ref-122)
123. - named entities [↑](#footnote-ref-123)
124. - Named Entity Recognition [↑](#footnote-ref-124)
125. - Wilson [↑](#footnote-ref-125)
126. - Xia [↑](#footnote-ref-126)
127. - Sparse Representation [↑](#footnote-ref-127)
128. - Dense Representation Based Method [↑](#footnote-ref-128)
129. - Das [↑](#footnote-ref-129)
130. - Gated Recurrent Unit [↑](#footnote-ref-130)
131. - Feldman [↑](#footnote-ref-131)
132. - Maximum Inner Product Search [↑](#footnote-ref-132)
133. - Seo [↑](#footnote-ref-133)
134. - Dohan [↑](#footnote-ref-134)
135. - Self\_attention [↑](#footnote-ref-135)
136. - Lee [↑](#footnote-ref-136)
137. - Guu [↑](#footnote-ref-137)
138. - masked language modeling [↑](#footnote-ref-138)
139. - Chen [↑](#footnote-ref-139)
140. - Document Postprocessing [↑](#footnote-ref-140)
141. - Answer Postprocessing [↑](#footnote-ref-141)
142. - pipeline systems [↑](#footnote-ref-142)
143. - End to End System [↑](#footnote-ref-143)
144. - Sparse Retriever [↑](#footnote-ref-144)
145. - Dense Retriever [↑](#footnote-ref-145)
146. - Itrative Retriver [↑](#footnote-ref-146)
147. - Dual-encoder [↑](#footnote-ref-147)
148. - Two-Tower [↑](#footnote-ref-148)
149. - Karpukhin [↑](#footnote-ref-149)
150. - Nishida [↑](#footnote-ref-150)
151. - Comperhension Layer [↑](#footnote-ref-151)
152. - Nie [↑](#footnote-ref-152)
153. - Zhao [↑](#footnote-ref-153)
154. - non-contextualized encoded [↑](#footnote-ref-154)
155. - contextualized encoded [↑](#footnote-ref-155)
156. - Multi-step Retriever [↑](#footnote-ref-156)
157. - multi-hope reasoning [↑](#footnote-ref-157)
158. - Qi [↑](#footnote-ref-158)
159. - Min [↑](#footnote-ref-159)
160. - Path Retriver [↑](#footnote-ref-160)
161. - Asai [↑](#footnote-ref-161)
162. - Latent Space [↑](#footnote-ref-162)
163. - Das [↑](#footnote-ref-163)
164. - Multi-step Reasoner [↑](#footnote-ref-164)
165. - Gated Recurrent Unit [↑](#footnote-ref-165)
166. - Kratzwald [↑](#footnote-ref-166)
167. - simple threshold based heuristic method [↑](#footnote-ref-167)
168. - ordinal ridge regression [↑](#footnote-ref-168)
169. - Path Retriver [↑](#footnote-ref-169)
170. - end\_of\_evidence [↑](#footnote-ref-170)
171. - Recurrent Retriver [↑](#footnote-ref-171)
172. - Precision [↑](#footnote-ref-172)
173. - Recall [↑](#footnote-ref-173)
174. - corpus [↑](#footnote-ref-174)
175. - Nishida [↑](#footnote-ref-175)
176. - Lin [↑](#footnote-ref-176)
177. - Schooten [↑](#footnote-ref-177)
178. - Relation-Networks Ranker [↑](#footnote-ref-178)
179. - Yun [↑](#footnote-ref-179)
180. - Sequence-to-Sequence [↑](#footnote-ref-180)
181. - Path Retriver [↑](#footnote-ref-181)
182. - Clark [↑](#footnote-ref-182)
183. - Tan [↑](#footnote-ref-183)
184. - Evidance Extraction [↑](#footnote-ref-184)
185. - Lewis [↑](#footnote-ref-185)
186. - Izacard [↑](#footnote-ref-186)
187. - Fusion-in-Decoder [↑](#footnote-ref-187)
188. - Wang [↑](#footnote-ref-188)
189. - strength-based re-ranker [↑](#footnote-ref-189)
190. - coverage-based re-ranker [↑](#footnote-ref-190)
191. - pseudo passage [↑](#footnote-ref-191)
192. - Feed-forward network [↑](#footnote-ref-192)
193. - Retriver - Reader [↑](#footnote-ref-193)
194. - Retriver - Only [↑](#footnote-ref-194)
195. - Retriver - Free [↑](#footnote-ref-195)
196. - Natural Language Generation [↑](#footnote-ref-196)
197. - Transformer-based architectures [↑](#footnote-ref-197)
198. - Radford [↑](#footnote-ref-198)
199. - Brown [↑](#footnote-ref-199)
200. - few-shot learning [↑](#footnote-ref-200)
201. - Petroni [↑](#footnote-ref-201)
202. - Norbert Wiener [↑](#footnote-ref-202)
203. - C. W. J. Granger [↑](#footnote-ref-203)
204. - Ishida [↑](#footnote-ref-204)
205. - Fukumoto [↑](#footnote-ref-205)
206. - non-factoid [↑](#footnote-ref-206)
207. - Rhetorical Structure theory [↑](#footnote-ref-207)
208. - Higashinaka [↑](#footnote-ref-208)
209. - Isozaki [↑](#footnote-ref-209)
210. - Maximum Inner Product Search [↑](#footnote-ref-210)
211. - Ferrucci [↑](#footnote-ref-211)
212. - Sun [↑](#footnote-ref-212)
213. - Min [↑](#footnote-ref-213)
214. - Black Box [↑](#footnote-ref-214)
215. - Conversational QA [↑](#footnote-ref-215)
216. - Question Generation [↑](#footnote-ref-216)
217. - Chen Qu [↑](#footnote-ref-217)
218. - Modular End to End System [↑](#footnote-ref-218)
219. - Hinton [↑](#footnote-ref-219)
220. - reconstruction error [↑](#footnote-ref-220)