
بسم الله الرحمن الرحيم

مرکز آموزشهای الکترونیکی
گروه مهندسی کامپیوتر

سیستم پاسخگویی تعاملی با استفاده از تکنیک‌های هوش مصنوعی

سلیمه سادات شهرآئینی

استاد راهنما
دکتر مرتضی زاهدی

استاد مشاور
مهندس مهدی یعقوبی

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

شهریور ۹۴

دانشگاه شاهرود

مرکز آموزشهای الکترونیکی
گروه مهندسی کامپیوتر

پایان نامه کارشناسی ارشد آقای/ خانم سلیمه سادات شهرآئینی به شماره دانشجویی: ۹۱۲۶۶۲۴
تحت عنوان: سیستم پاسخگویی تعاملی با استفاده از تکنیک‌های هوش مصنوعی

در تاریخ ..... توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد مورد ارزیابی و با درجه ..... مورد پذیرش قرار گرفت.

اساتید راهنما	امضاء	اساتید مشاور	امضاء
نام و نام خانوادگی :		نام و نام خانوادگی :	
نام و نام خانوادگی :		نام و نام خانوادگی :	

اساتید داور	امضاء	نماینده تحصیلات تکمیلی	امضاء
نام و نام خانوادگی :		نام و نام خانوادگی :	
نام و نام خانوادگی :			
نام و نام خانوادگی :			
نام و نام خانوادگی :			

بنام خدا

دیر زمانیست که توسن عقل را به تاخت در دشت بیکران دانش می‌رانم و هنوز امید رسیدنم نیست چه آنکه بزرگمرد دانش پارسی، خیام فرمود:

آنانکه محیط فضل و آداب شدند — در جمع کمال، شمع اصحاب شدند
ره زین شب تاریک نبردند به روز — گفتند فسانه‌ای و در خواب شدند

حمد و ستایش خدا راست در این کار و شکرانه استاد گرامی‌ام جناب آقای دکتر مرتضی زاهدی که پایمردانه در این تلاش یاری‌ام نمود و صد البته حسب حدیث شریف رضوی که:

من لم یشکر المنعم من المخلوقین لم یشکر الله عزوجل ... سخن کوتاه کرده و سر فرود می‌آورم به قدردانی از:

اساتید داور جناب آقای دکتر علی اکبر پویان و سرکار خانم دکتر سعیده فردوسی و استاد مشاور جناب آقای مهندس مهدی یعقوبی که با راهنمایی‌های ارزنده خویش مصباح هدایت این حقیر بودند.

صالح و طالح متاع خویش نمودند — تا چه قبول افتد و چه در نظر آید

تعهد نامه

اینجانب سلیمه سادات شهرآئینی دانشجوی دوره کارشناسی ارشد رشته مهندسی کامپیوتر-گرایش هوش مصنوعی مرکز آموزشهای الکترونیکی دانشگاه شاهرود نویسنده پایان نامه طراحی سیستم پاسخگویی تعاملی با استفاده از تکنیک‌های هوش مصنوعی تحت راهنمائی جناب آقای دکتر مرتضی زاهدی متعهد می‌شوم.

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه شاهرود می‌باشد و مقالات مستخرج با نام « دانشگاه شاهرود » و یا « Shahrood University» به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان نامه، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه های رایانه ای، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه شاهرود می باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

چکیده

در سیستم‌های پرسش و پاسخ اتوماتیک[1]، کاربران پرسش‌های خود را به زبان طبیعی مطرح می‌کنند و سیستم پاسخ ممکن را بازمی‌گرداند. با این وجود در سیستم‌های پرسش و پاسخ اتوماتیک، گفتگوی پیوسته بین کاربر و سیستم وجود ندارد و کاربران نمی‌توانند نیاز خود را توضیح دهند. از این‌رو دسترسی به اطلاعات مورد نیاز در اولین تعامل بین کاربر و سیستم بسیار مشکل است. افزودن تعامل به این سیستم‌ها، امکان طرح سوالات مرتبط[2] و ارائه توضیحات لازم از طریق گفتگو بین سیستم و کاربر را فراهم می‌کند و دقت پاسخگویی را افزایش می‌دهد.

هدف از این تحقیق، طراحی یک سیستم پرسش و پاسخ مبتنی بر تعامل[3] با استفاده از تکنیک‌های آماری جهت استخراج دانش نهفته در متون ساختارنیافته است. این سیستم برخلاف سیستم‌های پیشین که عمدتا از تجزیه و تحلیل‌های معنایی و گرامری استفاده می‌کنند، از روش‌های مبتنی بر معنا و گرامر استفاده نکرده و از این‌رو مستقل از زبان می‌باشد و با در اختیار داشتن پایگاه دادگان مناسب هر زبان می‌تواند به سوالات مطرح شده به آن زبان پاسخ دهد. سیستم طراحی شده از پایگاه دادگان فارسی که به این منظور طراحی شده است، جهت آموزش و ارزیابی استفاده می‌نماید. میزان کارایی سیستم طراحی شده با استفاده از اندازه‌گیری معیارهای کمی و کیفی سنجیده شده است. مقایسه مقادیر به دست آمده افزایش ۲۲,۳٪ دقت پاسخگویی را نسبت به سیستم‌های پرسش و پاسخ اتوماتیک نشان می‌دهد. همچنین بررسی نظرات ارائه شده توسط کاربران نشان‌دهنده رضایت آنها از کیفیت تعامل برقرار شده با سیستم می‌باشد.

کلمات کلیدی: سیستم پرسش و پاسخ تعاملی، بازیابی اطلاعات، سیستم پرسش و پاسخ، تعامل انسان و ماشین

[1] Question Answering (QA)

[2] Follow up question

[3] Interactive Question Answering (IQA)

لیست مقالات

[1] Shahraini S, Zahedi M, (2015), "A language-Independent Interactive Question Answering System", *International Journal* of Review in Life Sciences, 5(10), pp. 961-965.

[2] شهرآئینی س، زاهدی م، (۱۳۹۴) "مدل تعاملی جدید جهت استفاده در سیستم‌های پرسش و پاسخ اتوماتیک"، دومین کنگره بین المللی مهندسی برق، علوم کامپیوتر و فناوری اطلاعات، تهران

فهرست مطالب

لیست اختصارات

QA	Question Answering
IQA	Interactive Question Answering
NLP	Natural Language Proccesing
RDQA	Restricted Domain Question Answering
ODQA	Open Domain Question Answering
TREC	Text Retrieval Conference
CLEF	Cross Language Evolution Forum
NTCIR	NII Testbeds and Community for Information access Research
BOW	Bag-Of-Words
TF	Term Frequency
POS	Part-Of-Speech
IDF	Inverse Document Frequency
KNN	K-Nearest Neighbour
SVM	Support Vector Machine

فهرست شکل‌ها

فهرست جداول

فصل اول

مقدمه

۱-۱ مقدمه

در جهان پر از اطلاعات امروز، یافتن پاسخ‌های صحیح و دقیق برای سوالات موردنظر در کوتاه‌ترین زمان ممکن به یکی از چالش‌های افراد تبدیل شده است. برای پاسخگویی به این نیازمندی اطلاعاتی[۱] انواعی از سیستم‌های بازیابی اطلاعات طراحی شده‌اند که می‌توانند برای سوالات کاربران پاسخ‌هایی را به صورت متن، صوت، تصویر، ویدئو یا ترکیبی از آنها را ارائه کنند [۱]. با توجه به افزونگی اطلاعات[۲] در فضای وب بسیاری از افراد تلاش می‌کنند تا با بهره‌گیری از سیستم‌های بازیابی اطلاعات مانند سیستم‌های کلاسیک یاهو، گوگل و غیره نیازمندی‌های اطلاعاتی خود را در کوتاه‌ترین زمان ممکن برطرف نمایند. فرایند بازیابی اطلاعات توسط این سیستم‌ها شامل مراحل زیر است:

۱) کاربران با استفاده از کلمات و عبارات کلیدی نیاز اطلاعاتی خود را مطرح می‌کنند؛

۲) سیستم بازیابی اطلاعات اسناد مرتبط[۳] با نیاز اطلاعاتی کاربر را از بین مجموعه اسناد خود پیدا می‌کند؛

۳) سیستم اسناد یافت شده را بر اساس میزان ارتباطشان با پرسش مطرح شده توسط کاربر امتیازدهی[۴] می‌کند؛

۴) فهرستی از اسناد مرتبط که بیشترین امتیازها را توسط سیستم دریافت کرده‌اند، به عنوان پاسخ به کاربر ارائه می‌شود.

در این نوع بازیابی به جای پاسخ دقیق و مختصر، مجموعه‌ای از اسناد مرتبط به کاربر ارائه می‌شود و کاربر باید پاسخ موردنظر خود را از اسناد ارائه شده استخراج نماید که این فرایند نیازمند صرف زمان و تلاش قابل توجه از سوی کاربر می‌باشد. برای رفع این اشکال، سیستم‌های پرسش و پاسخ طراحی شده‌اند که قادرند سوال کاربر را به شکل زبان طبیعی دریافت کرده و پس از تجزیه و تحلیل سوال و

[1] Information need
[2] Redundancy
[3] Relevant
[4] Rank

نیز مجموعه اسناد مرتبط، پاسخ مناسب و دقیق را به صورت یک متن کوتاه و یا حتی یک جمله، عبارت یا کلمه بازگردانند.

پیدایش سیستم‌های پرسش و پاسخ به عنوان یکی از دستاوردهای علم رایانه در حوزه بازیابی اطلاعات و پردازش زبان طبیعی[1] به چند دهه قبل بازمی‌گردد [۱]. در این سیستم‌ها که شکل پیچیده‌تری از سیستم‌های بازیابی اطلاعات هستند، کاربر در هر بار نوبت‌گیری[2] فقط می‌تواند یک پرسش به صورت مستقل از پرسش‌های پیشین مطرح نماید تا سیستم با استفاده از تکنیک‌های موجود پاسخ احتمالی را به صورت کوتاه بازگرداند. پس از پایان فرایند پاسخگویی، سیستم پرسش مطرح شده و پاسخ آن را از حافظه خود پاک می‌کند. عدم وجود حافظه در این سیستم‌ها مطرح‌شدن مجموعه‌ای از سوالات مرتبط با یکدیگر را توسط کاربر غیر ممکن می‌سازد. علاوه بر این، موفقیت جستجو در این سیستم‌ها به نوع پرسش مطرح شده بستگی دارد. بعضی سوالات مطرح شده ساده بوده و پاسخگویی به آنها به راحتی انجام می‌شود. در مقابل سوالات پیچیده‌ای مطرح می‌شوند که پاسخگویی به آنها نیازمند طرح توضیحات اضافی از سوی کاربر می‌باشد. این محدودیت‌ها سبب می‌شود تا در بسیاری از موارد کاربران در یافتن پاسخ صحیح با مشکل مواجه شوند.

استفاده از مکالمات تعاملی در ارتباطات روزمره مردم را قادر می‌سازد که در صورت لزوم سخن خود را تصدیق یا اصلاح نمایند. این در حالیست که در سیستم‌های پرسش و پاسخ اتوماتیک امکان برقراری تعامل دو طرفه بین کاربر و سیستم وجود ندارد و کاربران قادر به ارائه توضیح اصلاحی درباره نیاز اطلاعاتی خود نیستند. بعلاوه سیستم نیز نمی‌تواند با طرح پرسش‌های خود از کاربر، جهت رفع ابهامات احتمالی در سوالات مطرح شده اقدام نماید. بدین ترتیب دسترسی به اطلاعات موردنظر در اولین پرسش و پاسخ چندان آسان نیست. در مقابل، سیستم‌های پرسش و پاسخ مبتنی بر تعامل امکان انجام مکالمات تعاملی و دو سویه بین کاربر و سیستم را فراهم نموده و رفع ابهام[3] از سوالات

[1] Natural Language Proccesing (NLP)

[2] Turn taking

[3] Disambiguation

مطرح شده توسط کاربر و تصحیح پاسخ‌های ارائه شده توسط سیستم را امکان‌پذیر می‌نمایند. همچنین در این سیستم‌ها تاریخچه گفتگو در حافظه نگهداشته می‌شود و بنابراین کاربر قادر خواهد بود مجموعه‌ای از سوالات مرتبط با یکدیگر را از سیستم بپرسد.

استفاده از سیستم‌های پرسش و پاسخ تعاملی جستجوی اطلاعات را تسهیل نموده و دقت پاسخگویی را افزایش می‌دهد. البته سیستم‌های پرسش و پاسخ تعاملی موجود عمدتا یا از تجزیه و تحلیل‌های معنایی و گرامری استفاده می‌کنند که آنها را محدود به زبان خاص می‌نماید و یا نیازمند آموزش با پیکره بزرگی از دادگان از طریق بکارگیری تکنیک‌های یادگیری ماشین می‌باشند. این محدودیت‌ها، طراحی یک سیستم پرسش و پاسخ مستقل از زبان را که بتواند از دانش نهفته در متون ساختارنیافته، اطلاعات مورد نیاز جهت برقراری تعامل با کاربران و برآوردن نیاز اطلاعاتی آنها را استخراج نماید ضروری می‌سازد.

۱-۲ هدف پایان‌نامه

هدف اصلی این پایان‌نامه، طراحی یک سیستم پرسش و پاسخ است که با استفاده از تکنیک‌های آماری بتواند دانش نهفته در متون ساختارنیافته را استخراج نماید و از آن جهت افزایش سطح تعامل با کاربر، مستقل از زبان و حوزه دانش استفاده نماید.

در این پایان‌نامه فرضیات زیر مطرح شده است:

۱) با استفاده از تکنیک‌های آماری می‌توان دانش نهفته در متون ساختارنیافته را استخراج نموده و از این دانش جهت افزایش سطح تعامل با کاربر، مستقل از زبان و حوزه دانش استفاده نمود؛

۲) بازیابی قطعه متون[1] به جای پاسخ دقیق، دامنه اطلاعات کاربر را افزایش داده و ممکن است او را از طرح سوالات مرتبط دیگر بی‌نیاز نماید.

به منظور دستیابی به هدف اصلی این پایان‌نامه مجموعه اهداف زیر تعریف شده‌اند:

[1] Passage retrieval

۱) مروری بر روش‌های موجود در طراحی سیستم‌های پرسش و پاسخ تعاملی

۲) تهیه و گردآوری پایگاه دادگان فارسی جهت ارزیابی سیستم‌های پرسش و پاسخ

۳) طراحی ماژول تعیین مراجع مرتبط با ضمایر[۱]

۴) طراحی ماژول بازسازی سوالات حذفی[۲]

۵) طراحی ماژول تصحیح خطای املایی[۳] و ابهام‌زدایی از لغات

۶) طراحی ماژول پیشنهاد دهنده پرسش‌های پرتکرار

۷) طراحی ماژول پیشنهاد دهنده عبارات اسمی در طرح پرسش

۸) طراحی ماژول بازیابی قطعه متون با تکنیک‌های آماری

۱-۳ حوزه

در روش پیشنهاد شده، هر یک از منابع داده و سوالات مطرح شده به عنوان مجموعه‌ای از کلمات شناخته می‌شوند که فارغ از معنا و گرامر بوده و سیستم تنها با استفاده از تکنیک‌های آماری محتمل‌ترین پاسخ را از بین گزینه‌های احتمالی انتخاب نموده و بازمی‌گرداند و محدودیتی در زمینه حجم منابع داده، حوزه دانش و زبان وجود ندارد. نکته قابل توجه دیگر این است که اسناد موجود در پایگاه دادگان باید فرمت متنی و یا XML داشته باشند.

از آنجا که تاکنون هیچ‌گونه دادگان فارسی جهت ارزیابی سیستم‌های پرسش و پاسخ فارسی ارائه نشده بود، در اولین گام از تحقیق سه پایگاه دادگان با مشخصات زیر جهت آموزش[۴] و ارزیابی[۵] سیستم‌های پرسش و پاسخ فارسی تهیه شده است:

[1] Anaphora resolusion
[2] Elliptic question
[3] Spell checking
[4] Train
[5] Test

۱. پایگاه دادگان اول با نام WMPR-QA1-2015 دارای چهار فایل متنی با محتوای آئین‌نامه آموزشی دانشگاه شاهرود می‌باشد که در قالب ۱۴۶ جمله و با فرمت UTF-8 گردآوری شده است و به عنوان داده آموزشی شناخته می‌شود. ۸۱ پرسش و پاسخ مطرح شده از این آئین‌نامه نیز به عنوان مجموعه تست پایگاه دادگان فوق شناخته می‌شود.

۲. پایگاه دادگان دوم با نام WMPR-QA2-2015 دارای یک فایل متنی با محتوای آئین‌نامه مالی شهرداری‌ها می‌باشد که در قالب ۷۵ جمله و با فرمت UTF-8 گردآوری شده است و از آن به عنوان مجموعه آموزش استفاده می‌شود. ۳۳ پرسش و پاسخ مطرح شده از این آئین‌نامه نیز به عنوان مجموعه تست پایگاه دادگان WMPR-QA2-2015 در نظر گرفته شده است.

۳. پایگاه دادگان سوم با نام WMPR-QA3-2015 شامل دو مجموعه آموزش و تست می‌باشد. مجموعه آموزش آن دارای یک فایل متنی با محتوای آئین‌نامه استخدام هیات علمی دانشگاه‌ها می‌باشد که در قالب ۲۵۶ جمله و با فرمت UTF-8 گردآوری شده است و مجموعه تست آن دربردارنده ۳۱ پرسش و پاسخ مطرح شده از این آئین‌نامه می‌باشد.

سه پایگاه دادگان فوق هم اکنون از آزمایشگاه وب‌کاوی و شناسایی الگو دانشگاه صنعتی شاهرود[1] قابل دریافت می‌باشند.

۱-۴ مروری برکارهای پیشین

یکی از نخستین سیستم‌های پرسش و پاسخ طراحی شده، سیستمی به نام ***BASEBALL*** بود که در سال ۱۹۶۱ جهت پاسخگویی به سوالات کاربران در رابطه با بازی‌های لیگ سالانه بیس‌بال در آمریکا طراحی شد [۱]. این برنامه رایانه‌ای قادر بود به سوالات انگلیسی کاربران در زمینه تاریخ و مکان برگزاری هر یک از بازی‌ها، تیم‌های شرکت‌کننده و امتیازات کسب شده در هر بازی پاسخ دهد. کاربران این سیستم سوال خود را به زبان انگلیسی و با استفاده از کارت‌پانچ‌ها مطرح می‌کردند. پس از

[1] http://wmpr.ir/fa/index/category/53

دریافت سوال، سیستم کلمات و اصطلاحات موجود در سوال را در یک لغت‌نامه جستجو نموده و با تحلیل ساختاری و گرامری پرسش، اطلاعات موجود در پرسش و نیز نوع اطلاعات مورد نیاز جهت ارائه پاسخ مناسب را تعیین می‌نمود. در مرحله بعد، سیستم اطلاعات موجود در پرسش را با اطلاعات موجود در پایگاه داده تطبیق داده و پس از استخراج پاسخ مناسب آن را به کاربر ارائه می‌کرد. علت موفقیت نسبی این سیستم، محدود بودن دامنه موضوع، ساده بودن داده و تکراری و آشنا بودن پرسش‌ها بوده است. در این سیستم که داده‌ها و لغت‌نامه دارای ساختار فهرست‌مانند بودند، عناصر اطلاعاتی به صورت زوج‌های مشخصه- مقدار[1] ذخیره می‌شدند [۲]. مثال زیر، فهرست مشخصات استخراج شده توسط سیستم از یک پرسش فرضی را ارائه می‌کند:

تیم ملی والیبال در چه تاریخی به برزیل سفر خواهد کرد؟ ⟵ مکان: برزیل
عامل: تیم ملی والیبال
تاریخ: ؟

نمونه دیگری از سیستم‌های پرسش و پاسخ به نام ***LUNAR*** در سال ۱۹۷۲ طراحی و ارائه گردید. هدف از طراحی این سیستم، پاسخگویی به سوالات زمین‌شناسان درباره نتایج حاصل از بررسی ساختار و ترکیب شیمیایی خاک و نمونه سنگ‌های کره ماه پس از پایان ماموریت سفینه آپولو بود [۱]. بررسی پاسخ‌های ارائه شده توسط این سیستم نشان داد که ۷۸٪ پاسخ‌ها صحیح بوده و فقط ۲۲٪ آنها دارای خطا بودند که از این میان ۱۲٪ خطاها به علت کدکردن نادرست لغت‌نامه و تنها ۱۰٪ آنها در نتیجه تجزیه و تحلیل نادرست معنایی بوده و خطای جدی محسوب می‌شدند [۳].

مهمترین ویژگی‌های عملکردی این دو سیستم و سیستم‌های مشابه آنها عبارتند از [۴]:

۱) تولید پایگاه داده‌های ساخت‌یافته (پایگاه دانش) از اطلاعات

۲) ساخت مجموعه عظیمی از الگوهای تحلیلی با دامنه مشخصی از واژگان به صورت دستی

۳) تجزیه و تحلیل پرسش‌های دریافت شده با استفاده از مجموعه الگوهای موجود

[1] Attribute-Value

۴) استفاده از الگو برای ترجمه پرسش زبان طبیعی به شکل یک پرس‌وجوی ساخت‌یافته

۵) جستجوی پرسش ساخت‌یافته در پایگاه داده ساخت‌یافته

در دهه ۱۹۷۰، نوع دیگری از سیستم‌های پرسش و پاسخ که استدلال محور بوده و قابلیت پاسخ به سوالات کاربر در یک دامنه محدود را داشتند ارائه گردیدند. این‌گونه سیستم‌ها قادر بودند با کاربر تعامل برقرار نموده و برای درک بهتر اهداف او از وی سوالاتی بپرسند و سپس با استفاده از دانش موجود در پایگاه داده و نیز اطلاعات تکمیلی ارائه شده توسط کاربر، به سوالات او پاسخ دهند [۱]. از جمله این سیستم‌ها می‌توان به سیستم *SHRDLU* در سال ۱۹۷۲ اشاره نمود که همانند یک ربات شبیه‌سازی شده عمل می‌کرد و می‌توانست ضمن پاسخگویی به سوالات کاربر درباره محیط اطراف خود، اشیا و بلوک‌های اسباب‌بازی را بر طبق خواست کاربر جابجا نماید. این سیستم که با مجموعه‌ای از قوانین ساده فیزیک در رایانه کد شده بود، به وسیله وینوگراد[1] و با استفاده از زبان برنامه‌نویسی LISP طراحی شده بود [۱]. در این سیستم فقط از ۵۰ کلمه جهت مشخص نمودن اشیا و مکان‌های موردنظر استفاده می‌شد. به دلیل دامنه محدود لغات استفاده شده، درک منظور کاربر برای سیستم بسیار ساده بود. این سیستم از یک حافظه کوچک جهت ثبت دستورات گذشته کاربر استفاده می‌کرد و به وسیله آن می‌توانست مرجع ضمایر بکار رفته در جمله کنونی کاربر را در سابقه دستورات قبلی او پیدا نماید[2]. همچنین وجود این حافظه سیستم را قادر می‌ساخت اسامی جدید اشیا را که توسط کاربران به کار می‌رفتند به خاطر سپرده و به این ترتیب دامنه لغات شناخته شده برای خود را افزایش دهد.

EAGLi، نام سیستم پرسش و پاسخ دیگری است که هدف آن پاسخگویی به سوالات کاربر در حوزه پزشکی بود. این سیستم از یک کتابخانه دیجیتال به نام MEDLINE که بیش از ۱۸ میلیون مقاله را در خود جای داده بود و هر سال ۸۰۰۰۰۰ مقاله به آن اضافه می‌گردید و یک موتور جستجو به نام PubMed بهره می‌برد. ارزیابی عملکرد این سیستم نشان داد که ۵۷٪ پاسخ‌های داده شده به سوالات

[1] Winograd

[2] Coreference

مربوط به بیماری‌ها و پروتئین‌ها و ۶۸٪ پاسخ‌های مربوط به بیماری‌ها و داروها صحیح بودند. ۷۰٪ از پاسخ‌های مورد اول و ۷۵٪ از پاسخ‌های مورد دوم جز ۱۰ پاسخ محتمل ارائه شده در ابتدای فهرست پیشنهادی سیستم قرار داشتند [۵]. علاوه بر سیستم فوق، سیستم‌های پزشکی دیگری مانند MYCIN نیز برای ارائه پاسخ به سوالات پزشکی کاربر طراحی شده‌اند [۱].

این سیستم‌ها و سیستم‌های دیگری مانند LIFER، QUALM، CHAT80، LADDER، PLANES وTEAM که برای پاسخ به سوالات کاربر در یک حوزه خاص طراحی شده‌اند، **سیستم‌های پرسش و پاسخ با دامنه محدود**[1] و گاهی سیستم‌های ایستا[2] نامیده می‌شوند. این سیستم‌ها از منابع دانش یا پایگاه داده‌هایی که به صورت دستی طراحی می‌شوند استفاده نموده و دامنه مشخص و محدودی از اصطلاحات و واژگان پرکاربرد در پرسش‌های کاربر در آن حوزه خاص را دارا می‌باشند [۶].

با برگزاری کنفرانس سالانه بازیابی متون (TREC)[3] در سال ۱۹۹۹، دوره جدیدی در طراحی سیستم‌های پرسش و پاسخ آغاز گردید [۷،۱]. در این سال، کنفرانس مذکور (TREC-8) با اعلام رویکرد جدید خود طراحان سیستم‌های پرسش و پاسخ را به طراحی سیستم‌هایی تشویق نمود که با بهره‌گیری از مجموعه بزرگی از اسناد متنی در موضوعات مختلف، بتوانند پاسخ کوتاهی برای سوال کاربر که درباره موضوعات مختلف با زبان طبیعی مطرح شده است ارائه نمایند. این‌گونه سیستم‌ها **سیستم‌های با دامنه نامحدود**[4] و گاهی سیستم‌های پویا[5] نامیده می‌شوند [۶].

اگرچه جریان اصلی طراحی سیستم‌های دامنه نامحدود پس از کنفرانس TREC-8 در سال ۱۹۹۹ آغاز شد، اولین سیستم پرسش و پاسخ مبتنی بر وب با نام ***START*** در سال ۱۹۹۳ و در موسسه

[1] Restricted Domain Question Answering systems (RDQA)

[2] Static

[3] Text REtrieval Conference

[4] Open Domain Question Answering systems (ODQA)

[5] Dynamic

فناوری ماساچوست آمریکا طراحی گردید [۱۰،۹،۸،۱]. در این سیستم، از یک مدل داده‌ای ۳-گانه[1] بصورت "شی-ویژگی-مقدار"[2] جهت ساماندهی و یکپارچه‌سازی منابع اطلاعاتی ساخت‌یافته و غیرساخت‌یافته ناهمگون وب استفاده شده است [۱۰،۸]. استفاده از این مدل داده‌ای باعث می‌شود مفاهیم مختلف موجود در جملات و عبارات زبان طبیعی به فرمتی قابل درک برای رایانه تبدیل شوند. به همین دلیل قبل از الحاق منبع اطلاعاتی جدید به پایگاه داده مورد استفاده در START، منبع اطلاعاتی باید حاشیه‌نویسی[3] شود که معمولا این امر به صورت دستی انجام می‌شود، اگرچه محققان تلاش‌هایی را جهت استفاده از روش‌های متفاوتی برای خودکار کردن فرایند تولید حاشیه‌ها و تفاسیر انجام داده‌اند [۸].

در فرایند یافتن پاسخ مناسب برای سوال کاربر، این سیستم و سیستم‌های مشابه با دو چالش درک پرسش و یافتن محل جستجوی آن و ارائه پاسخ روبرو می‌باشند. باید توجه داشت که سیستم START ابتدا سوال کاربر را به تفاسیر ۳-گانه مربوطه تبدیل نموده و سپس تفاسیر ۳-گانه را به Omnibase که به عنوان پایگاه مجازی دادگان[4] و واسطی جهت دسترسی به منابع دانش چندگانه استخراج شده از وب عمل می‌کند، تحویل می‌دهد. هدف اصلی سیستم START ارائه پاسخ با دقت بالا به کاربر می‌باشد [۱۰].

کنفرانس ارزیابی بین‌زبانی که به اختصار آن را CLEF[5] می‌نامند در سال ۲۰۰۰ مفهوم پرسش و پاسخ بین‌زبانی[6] را مطرح نمود که به کاربر امکان می‌دهد سوال خود را به زبانی غیر از زبان اصلی اسناد موجود در پایگاه اطلاعات مطرح نماید. در این حالت سیستم باید بتواند سوال کاربر را ترجمه نموده و پاسخ سوال ترجمه شده را در میان اطلاعات جستجو نماید و یا اینکه بخش‌های کوچکی از

[1] Ternary Expression = T-Expression (TE)
[2] Object-Property-Value
[3] Annotation
[4] Virtual database
[5] Cross Language Evolution Forum
[6] Cross language question answering (Multilingual)

اسناد را که احتمالا با پرسش اصلی کاربر مطابقت دارند ترجمه نموده و به جستجوی پاسخ در اسناد ترجمه شده اقدام نماید [۱].

پرسش و پاسخ بین‌رسانه‌ای[1] مفهوم دیگری است که با الهام از مفهوم پرسش و پاسخ بین زبانی مطرح شده است و بر اساس آن سیستم پاسخ پرسش کاربر را که به زبان طبیعی مطرح شده است، در منابع اطلاعاتی که به شکل تصویر، ویدئو و یا صوت می‌باشند جستجو می‌نماید. مفهوم دیگری که توسط NTCIR[2] اخیرا توسعه داده شده است، پرسش و پاسخ زمانی-جغرافیایی می‌باشد که بر طبق آن بازیابی اطلاعات از منابع داده با در نظر گرفتن جنبه‌های زمانی و مکانی پرسش صورت می‌پذیرد [۱].

همچنین طراحی سیستم‌های پرسش و پاسخ تعاملی که بتواند با ایجاد تعامل با کاربر نسبت به رفع ابهام از سوال وی اقدام نماید، بخش مهمی از فعالیت‌های اخیر محققان را به خود اختصاص داده است. به عنوان مثال می‌توان به سیستم‌های پرسش و پاسخ تعاملی زیر اشاره نمود:

اولین سیستم‌های محاوره‌ای نظیر ELIZA [۱۱] و GUS [۱۲] از یک پایگاه داده خاص و ساختاریافته به عنوان منبع دانش استفاده می‌کردند. این سیستم‌ها فقط قابلیت پاسخگویی به پرسش‌های مطرح شده در زمینه‌های محدود را داشتند. سپس سیستم‌های تعاملی دیگری با افزودن مدیریت محدود به سیستم QA طراحی شدند. این سیستم‌ها محدودیت‌ها در سوال را شناسایی کرده و در صورتی‌که نیاز باشد برای اصلاح آنها با کاربر وارد تعامل می‌شوند. از جمله این سیستم‌ها می‌توان به سیستمی اشاره کرد که توسط کیو و گرین [۱۳] جهت دسترسی به اطلاعات پرواز طراحی شده است. همچنین رایزر و لمون [۱۴] در سال ۲۰۰۹ سیستمی را با بکارگیری روش یادگیری تقویتی ارائه نمودند که محدود به دامنه خاص می‌باشد. این سیستم‌ها Frame-based بوده و چنانچه پس از مقداردهی فیلدهای مربوطه توسط کاربر پاسخ مناسبی یافت نشود، سیستم با هدف درخواست تغییر سوال یا توضیح با کاربر وارد تعامل می‌شود. دورنسکیو و اوراسان [۱۵] سیستم پرسش و پاسخ تعاملی

[1] Cross-media

[2] NII Testbeds and Community for Information access Research

در چهارچوب (QALL-ME) طراحی کردند که جزء سیستم‌های QA دامنه محدود محسوب می‌شود و از تکنیک‌های پیچیده پردازش زبان طبیعی استفاده می‌کند. کوارترونی و ماناندار [۱۶] یک سیستم IQA پیشنهاد کردند که ترکیبی از سیستم QA دامنه باز و چت‌بات[1] است. این سیستم از زبان AIML و رویکرد تطابق الگو استفاده می‌کند. محققی به نام تانگ [۱۷] یک سیستم پرسش و پاسخ تعاملی جهت پاسخگویی به سوالات چینی طراحی کرده است که از شبکه معنایی[2] و پارسرهای موجود در زبان چینی استفاده می‌کند. کنستانتینوا [۱۸] سیستم پرسش و پاسخی محدود به دامنه خاص و دارای منابع متنی نیمه ساخت‌یافته طراحی کرده است که از طریق برقراری تعامل و اطلاع از نیازمندی کاربران، به آن‌ها در انتخاب و خرید تلفن همراه مورد نظرشان کمک می‌کند.

سیستم‌های موجود صرف نظر از حوزه دانش خود (محدود به دامنه دانش خاص و یا دامنه نامحدود)، عمدتا از تجزیه و تحلیل‌های معنایی و گرامری، تطابق الگو و یا آموزش با استفاده از مجموعه بزرگی از دادگان استفاده می‌کنند. تعامل این سیستم‌ها با کاربر با هدف بهینه‌سازی پرسش و رفع ابهام از آن و یا طرح پرسش‌های مرتبط انجام می‌شود.

۱-۵ ساختار پایان نامه

این پایان‌نامه از ۶ فصل تشکیل شده است. فصل اول خواننده را با مفهوم تعامل و بکارگیری آن در سیستم‌های پرسش و پاسخ آشنا می‌سازد و کارهای انجام شده در این حوزه را بررسی می‌کند. در فصل دوم، توضیحاتی درباره الگوریتم‌های پرکاربرد در حوزه‌های پرسش و پاسخ، بازیابی اطلاعات و پردازش زبان طبیعی با هدف شناخت مزایا، معایب و کاربرد هر یک از آنها ارائه شده است. فصل سوم به بررسی سیستم‌های پرسش و پاسخ، طبقه‌بندی‌های موجود، اجزای تشکیل دهنده و روش‌های ارزیابی آن‌ها می‌پردازد. فصل چهارم با معرفی سیستم‌های پرسش و پاسخ تعاملی، روش‌های موجود

[1] Chatbot
[2] Wordnet

جهت طراحی این سیستم‌ها و ارزیابی آن‌ها را معرفی می‌نماید. فصل پنجم ابتدا روش پیشنهادی جهت طراحی سیستم پرسش و پاسخ تعاملی مستقل از زبان و گام‌های مورد نیاز جهت پیاده‌سازی آن را معرفی می‌کند و سپس نتایج آزمایشات انجام شده بر روی سیستم طراحی شده را ارائه می‌نماید. فصل ششم به جمع‌بندی مطالب فصل‌های پیشین و ارائه پیشنهاداتی درباره چگونگی بهبود سیستم طراحی شده به عنوان کارهای آینده می‌پردازد.

فصل دوم

معرفی متدهای پرکاربرد در حوزه پرسش و پاسخ

۲-۱ مقدمه

هدف از طراحی سیستم پرسش و پاسخ، یافتن کوتاهترین و دقیق‌ترین پاسخ احتمالی برای پرسش کاربر است. سیستم پرسش و پاسخ از بخش‌ها و مراحل مختلفی تشکیل شده است که هر یک وظیفه خاصی را بر عهده دارند. روش‌های متعدد و متفاوتی جهت طراحی و پیاده‌سازی هر یک از این بخش‌ها ارائه شده است. برخی از سیستم‌های پرسش و پاسخ از ترکیب تکنیک‌های مختلف جهت افزایش دقت پاسخگویی و بهبود نتایج استفاده می‌کنند. در این فصل روش‌های پرکاربرد در طراحی این سیستم‌ها با هدف شناخت مزایا، معایب و کاربرد هریک از آن‌ها تشریح می‌شود.

۲-۲ طبقه بندی الگوریتم‌های موجود

سیستم‌های پرسش و پاسخ از الگوریتم‌های مختلفی برای پردازش و پاسخگویی به سوالات کاربران استفاده می‌کنند. در یک طبقه‌بندی کلی، الگوریتم‌های موجود در سه دسته قرار می‌گیرند [۲۰،۱۹]:

۱) روش‌های زبان‌شناختی: با توجه به اهمیت درک متون نوشته شده به زبان طبیعی در سیستم‌های پرسش و پاسخ، محققان در این روش با ترکیب تکنیک‌های پردازش زبان طبیعی و ساخت پایگاه‌های دانش به طراحی سیستم‌های پرسش و پاسخ کارآمد و دقیق می‌پردازند. در این سیستم‌ها با تجزیه و تحلیل‌های دقیق معنایی و گرامری، پرسش کاربر به یک پرس‌وجوی ساخت‌یافته تبدیل می‌شود و در صورت برخورداری سیستم از یک پایگاه دانش ساخت‌یافته، امکان دستیابی به پاسخ‌های دقیق، کوتاه و قابل اعتماد فراهم می‌شود. تهیه پایگاه دانش ساخت‌یافته، زمان‌بر بوده و کاربرد سیستم را نیز به دامنه‌های موضوعی خاص محدود می‌کند.

۲) روش‌های آماری: رشد سریع پایگاه‌های داده مبتنی بر وب، نیاز به بکارگیری روش‌های آماری را افزایش داده است، زیرا این روش‌ها با حجم زیاد داده‌ها و نیز غیرساخت‌یافته بودن منابع داده وب سازگارند و در نتیجه می‌توانند پاسخگوی پرسش‌های ساخت‌نیافته نیز باشند. آنچه در این سیستم‌ها

مورد توجه است ضرورت آموزش آنها می‌باشد که باید با استفاده از حجم بزرگی از داده‌های آموزشی انجام شود. این روش‌ها تنها در صورت آموزش صحیح، توانایی پاسخگویی دقیق خواهند داشت.

۳) روش‌های تطابق الگو: در این روش قوانین توسط افراد خبره و در حوزه خاصی از دانش تعریف می‌شود. رویکرد تطابق الگو یکی از تکنیک‌های ساده، مبتنی بر دانش و بدون استفاده از روش‌های پیچیده پردازش زبان طبیعی است که می‌تواند در سیستم‌های پرسش و پاسخ جهت انتخاب مناسب‌ترین پاسخ استفاده شود. در این روش با مقایسه ورودی و الگوهای از پیش تعریف شده، شبیه‌ترین الگو تعیین و با توجه به آن، مناسب‌ترین پاسخ انتخاب می‌شود. مقایسه سه روش فوق در جدول ۲-۱ آمده است:

جدول ۲-۱ مقایسه کلی روش‌های بکار رفته در سیستم‌های پرسش و پاسخ [۲۰]

	زبان شناختی	**آماری**	**تطابق الگو**
نوع پرسش	پرسش‌های حقیقی	پرسش‌های پیچیده و حقیقی	پرسش‌های حقیقی، تعریفی، شکل کامل کلمات مخفف
درک معنایی	زیاد	کم	کمتر از دو روش دیگر
مدیریت داده‌های ساخت‌نیافته	مشکل است، زیرا پایگاه‌های دانش تنها برای مدیریت داده‌های از پیش تعریف شده طراحی شده‌اند.	به این منظور از اندازه‌گیری شباهت آماری استفاده می‌شود.	به راحتی امکان‌پذیر است.
قابلیت اعتماد	زیاد	به دلیل بکارگیری روشهای یادگیری باناظر قابل اعتماد است.	وابسته به اعتبار پایگاه دانش است.
مقیاس پذیری	کم است. چون برای هر مفهوم جدید باید قوانین جدید به پایگاه دانش افزوده شوند.	اگر آموزش صحیح ببیند به برای داده های زیاد نیز کاربرد خواهد داشت.	کم است و برای هر مفهوم جدید، الگوهای جدید باید فرا گرفته شوند.
حوزه کاربرد	دامنه خاص	حجم زیاد داده مانند وب	وب‌سایت‌های کوچک

۲-۳ مدل‌های بازیابی اطلاعات و الگوریتم‌های رتبه‌بندی

در اولین گام برای طراحی سیستم‌های بازیابی اطلاعات، مدلی برای توصیف و تعیین ارتباط میان اطلاعات سیستم و نیازهای اطلاعاتی کاربر انتخاب می‌شود. تفاوت کارایی سیستم‌های پرسش و پاسخ مختلف ناشی از الگوریتم‌ها و مدل‌های مختلفی است که در این سیستم‌ها پیاده‌سازی می‌شوند. این مدل‌ها با توجه به مجموعه داده‌های مورد استفاده و زبان مقصد کارایی متفاوتی دارند. در این بخش مدل‌های مورد استفاده در سیستم‌های پرسش و پاسخ متنی برای بازیابی اطلاعات و نیز رتبه‌بندی آن‌ها بیان می‌شود.

۲-۳-۱ مدل کیسه‌ای از کلمات

مدل کیسه‌ای از کلمات[1] ساده‌ترین روش جهت بازیابی اطلاعات است که در آن اطلاعات موجود در پرسش کاربر و مجموعه اسناد متنی، تنها به عنوان مجموعه‌ای از کلمات شناخته می‌شوند. البته باید در نظر داشت که در این مدل، کلمات عمومی[2] و فاقد ارزش معنایی از مجموعه کلمات فوق حذف می‌شوند. همچنین در این مدل، ویژگی‌های گرامری، ارتباط معنایی بین لغات جمله و نیز موقعیت کلمات و ترتیب قرار گرفتن آن‌ها در جمله اهمیتی ندارد و نادیده گرفته می‌شود [۲۱،۱۹]. مراحل بازیابی اطلاعات در مدل کیسه‌ای از کلمات به شرح زیر می‌باشد [۱۹،۴]:

۱) از مجموعه اسناد موجود، کلیه کلمات به جز کلمات عمومی و فاقد ارزش استخراج و در یک لغت‌نامه جمع‌آوری می‌شود؛

۲) ماتریسی تعریف می‌شود که در آن هر سطر معادل یک لغت از لغت‌نامه و هر ستون معادل نام یکی از اسناد است. همچنین هر درایه از ماتریس مقدار صفر یا یک می‌پذیرد که نشان می‌دهد هر

[1] Bag-Of-Words (BOW)

[2] Stopword

کلمه در کدام سند یا اسناد موجود است. در این بازیابی، کلمات وزن‌دهی[1] نمی‌شوند و در نتیجه تفاوتی بین کلمات وجود ندارد.

۳) با بکارگیری یک تابع بازیابی که از مدل دودویی[2] استفاده می‌کند، سیستم پرسش و پاسخ مجموعه‌ای از اسناد مرتبط با پرس‌وجوی کاربر را به وی بازمی‌گرداند.

با وجود سادگی، این مدل معایب زیر را دارد:

- تبدیل متون و پرسش کاربر به یک مدل دودویی مستلزم صرف زمان و حافظه زیادی است.
- با توجه به اینکه این روش تنها به وجود یا عدم وجود کلمات پرسش در اسناد اهمیت می‌دهد، ممکن است تعداد اسناد بازیابی شده بسیار کمتر یا بیشتر از حد معمول باشد.
- جملاتی وجود دارند که دارای کلمات مشابه هستند، اما ساختار و معنای متفاوتی دارند. از آنجا که مدل کیسه‌ای تمایزی بین این جملات قائل نمی‌شود، سیستم پرسش و پاسخی که از این مدل استفاده کند ممکن است پاسخ نادرستی به کاربر ارائه نماید.

۲-۳-۲ مدل N-گرام

یک N-گرام دنباله‌ای از n کلمه یا n واج است که به صورت متوالی در یک متن ظاهر شده‌اند. جملاتی با موضوعات یکسان، عمدتا از N-گرام‌های مشابه استفاده می‌کنند و بنابراین با محاسبه تعداد تکرار N-گرام‌های عبارت کاربر در هر یک از متون پایگاه داده، میزان ارتباط آن متن با عبارت مطرح شده توسط کاربر مشخص می‌شود. N-گرام با اندازه یک، 1-گرام[3] و با اندازه‌های دو و سه به ترتیب ۲-گرام[4] و ۳-گرام[5] نامیده می‌شوند [۲۲]. مدل آماری N-گرام از سایر روش‌ها عملی‌تر و موثرتر می‌باشد زیرا علاوه بر سادگی پیاده‌سازی، نسبت به نویز نیز مقاوم است.

---

[1] Term weighting
[2] Boolean model
[3] Unigram
[4] Bigram
[5] Trigram

از ویژگی‌های قابل انتظار مدل‌های N-گرام این است که دقت و کارایی مدل با افزایش مقدار N افزایش می‌یابد. با وجود این واقعیت در بیشتر کاربردها عملا از مدل‌های ۲-گرام یا حداکثر ۳-گرام استفاده می‌شود، زیرا مدل‌های مرتبه بالاتر از ۳ برای آموزش مناسب احتیاج به مجموعه متون بزرگتری دارند و در غیر این‌صورت نمی‌توان تخمین‌های مناسبی برای احتمالات به دست آورد. ویژگی دیگر مدل‌های N-گرام وابستگی زیاد آن‌ها به متون آموزشی از نظر نوع و اندازه می‌باشد. اگر مجموعه آموزشی تنها مربوط به یک زمینه و موضوع خاص باشد، احتمالات نتیجه شده نمی‌توانند برای جملات جدید به شکلی مناسب تعمیم یابند. از طرفی اگر مجموعه متون آموزشی بسیار عمومی و کلی باشد، ممکن است احتمالات نتیجه شده برای آن کاربرد خاص مناسب نباشند [۲۳،۲۲].

۲-۳-۳ مدل فضای برداری

یکی از مدل‌های بازیابی اطلاعات مدل فضای برداری[1] می‌باشد که در سطح گسترده‌ای به کار می‌رود. در این مدل پرسش کاربر و هر کدام از اسناد به برداری از کلمات تبدیل می‌شوند. زاویه بین بردار هر سند و پرسش مطرح شده میزان شباهت آن سند را با پرسش کاربر نشان می‌دهد. مراحل بازیابی اطلاعات در این مدل به شرح زیر می‌باشد:

۱) لغت‌نامه‌ای شامل تمام لغات موجود در مجموعه اسناد و پرسش تشکیل می‌شود. برای کاهش اندازه لغت‌نامه عمدتا از ریشه لغات استفاده می‌شود و واژه‌های پرتکرار که فاقد بار معنایی هستند از این مجموعه حذف می‌شوند.

۲) پس از ساخت لغت‌نامه، متناظر با پرسش و هر یک از اسناد موجود برداری تعریف می‌شود که طول آن برابر طول لغت‌نامه می‌باشد و هر درایه از آن متناظر با یک لغت از لغت‌نامه است. مقدار هر درایه بیانگر اهمیت آن لغت می‌باشد. ساده‌ترین روش برای مقداردهی درایه‌ها این است که اگر لغت متناظر در پرسش و یا سند مربوطه وجود داشت به آن درایه مقدار یک و در غیر اینصورت مقدار

[1] Algebraic vector space model

صفر اختصاص یابد. روش دیگر این است که تعداد دفعات تکرار آن لغت در پرسش و یا سند[1] مورد نظر، در درایه ذخیره شود. در پرکاربردترین روش که TF*IDF نام دارد، وزن هر واژه با توجه به معکوس فراوانی رخداد اصطلاح در کل مجموعه اسناد و نیز تعداد دفعات حضور یک اصطلاح در سند خاص تعیین می‌شود.

۳) برای تعیین میزان شباهت پرسش با هر یک از اسناد، کافی است میزان شباهت بردارهای پرسش و سند مربوطه را با استفاده از روش‌های تعیین شباهت دو بردار، مانند روش مشابهت کسینوسی محاسبه نمود [۲۲،۱].

یکی از مهم‌ترین اشکالات این روش این است که لغات مستقل از یکدیگر در نظر گرفته شده و وابستگی بین آن‌ها نادیده گرفته می‌شود که این امر می‌تواند به کاهش کارایی سیستم منجر شود. مزیت این مدل این است که درجه ارتباط پرسش با هر یک از اسناد را نشان می‌دهد [۲۲].

۲-۳-۴ مدل احتمالاتی

مدل احتمالاتی[2] یکی از مدل‌های بازیابی اطلاعات است که در آن پس از طرح پرسش از سوی کاربر، احتمال ارتباط کلیه اسناد با نیاز اطلاعاتی محاسبه می‌شود و در نهایت فهرست مرتب شده‌ای از اسناد به صورت درجه‌بندی شده به کاربر ارائه می‌شود. بدین ترتیب اولین سند موجود در فهرست، محتمل‌ترین سند از نظر دارا بودن پاسخ سوال کاربر می‌باشد. در این روش یک مدل زبان برای سوال و یک مدل زبان برای سند در نظر گرفته شده و روابط زیر تعریف می‌شوند:

$$P(Q|D) = \prod_{i=1...M} P(q_i|D) \qquad (۱-۲)$$

P(D) احتمال اولیه سند و P(Q|D) احتمال شرطی وجود سوال در یک سند می‌باشد.

[1] Term Frequency (TF)

[2] Probabilistic model

فرمول فوق با فرض استقلال هر یک از کلمات کلیدی سوال که با q_i نمایش داده می‌شوند، محاسبه می‌گردد و M بیانگر تعداد کلمات کلیدی در سوال است. آنگاه با توجه به فرمول بیز، احتمال مرتبط بودن هر سند با سوال محاسبه شده و بر این اساس اسناد مرتب می‌شوند:

$$P(D|Q) \propto P(Q|D)P(D) \quad (۲-۲)$$

این مدل اسناد و پرسش‌ها را به برداری از کلمات تبدیل می‌کند و از این نظر شبیه مدل برداری عمل می‌نماید. البته مدل احتمالاتی برخلاف مدل برداری اسناد را نه بر اساس میزان مشابهت آنها با پرسش، بلکه بر اساس میزان احتمال ارتباطشان با پرسش بازیابی و مرتب می‌کند.

لازم به ذکر است که مدل‌های احتمالاتی هم می‌توانند از ویژگی‌های کاراکتری و هم از ویژگی‌های لغوی استفاده نمایند. یکی از نمونه‌های مدل احتمالاتی، روش بیزین است که معروفیت زیادی داشته و هموارسازی[1] را به خوبی فراهم می‌کند [۲۲].

باید توجه داشت که در مدل احتمالاتی، ارتباط بین لغات مانند مترادف بودن نادیده گرفته می‌شود و کلمات نیز مستقل از یکدیگر فرض می‌شوند. مرتبه زمانی بالا و پیچیدگی زیاد پیاده‌سازی نیز، مقیاس‌پذیر[2] بودن این مدل را دشوار ساخته است [۲۲].

۲-۳-۵ تجزیه و تحلیل ساختاری،گرامری و معنایی

گرامر یک زبان، مجموعه‌ای از دستورات زبانی است که با استفاده از آن می‌توان یک جمله را به اجزای نحوی آن تفکیک نمود. تجزیه و تحلیل جمله و شکستن آن به اجزای تشکیل دهنده مانند اسم، فعل، صفت، قید و مضاف‌الیه توسط پارسرهای زبان انجام می‌شود. باید توجه داشت که نادیده گرفتن ساختار جملات و نقش معنایی لغات می‌تواند فهم جمله را دچار مشکل نماید [۱]. به عنوان

[1] Smoothing
[2] Scalable

مثال، روش کیسه‌ای از کلمات و دیگر روش‌های مشابه که بدون توجه به ساختار جمله تنها به کلمات اهمیت می‌دهند تمایزی بین دو جمله زیر قائل نمی‌شوند.

الف) رئیس‌جمهور از کدام سفر برگشت؟

ب) کدام رئیس‌جمهور از سفر برگشت؟

در حالی‌که با در نظر گرفتن ساختارگرامری، تفاوت دو جمله فوق از یکدیگر تشخیص داده می‌شود. برخی از اقداماتی که در تجزیه و تحلیل ساختار گرامری جملات انجام می‌شود، عبارتند از:

۱) ریشه‌یابی کلمات[1]: ریشه یک کلمه با حذف پیشوندها و پسوندها در آن کلمه به دست می‌آید. جهت همسان‌سازی متون ابتدا با استفاده از ریشه‌یابی، ریشه لغات تعیین شده و سپس هر لغت با ریشه آن جایگزین می‌شود. بنابراین کلیه لغات دارای ریشه یکسان، به فرمت یکسانی تبدیل خواهند شد. الگوریتم‌های مختلفی جهت ریشه‌یابی وجود دارند که عمدتا مبتنی بر قانون هستند [۲۴،۱]؛

۲) شناسایی و برچسب زدن اجزای واژگانی کلام[2] : برچسب‌گذاری اجزای کلام، به معنای اختصاص برچسب‌های واژگانی به کلمات تشکیل دهنده یک متن است و این برچسب‌ها نشان‌دهنده نقش کلمات در جمله می‌باشند. کلمات بکار رفته در جایگاه‌های مختلف نقش‌های متفاوتی دارند. برچسب‌گذاری واژگانی در حوزه‌های مختلف پردازش زبان طبیعی مانند ترجمه ماشینی، خطایابی و تبدیل متن به گفتار نقش مهمی دارد [۲۵،۱]. جدول۲-۲ نمونه‌ای از یک مجموعه برچسب[3] برای لغات را نشان می‌دهد.

[1] Stemming

[2] Part-Of-Speech (POS)

[3] Tagset

جدول ۲-۲ مجموعه برچسب‌های لغات

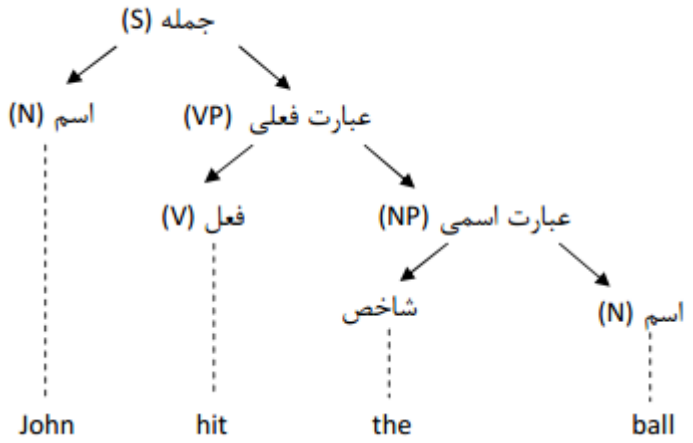
ADV	Adverb	قید
AJ	Adjective	صفت
CL	Classifier	شاخص
CONJ	Conjunction	حرف ربط
DET	Determiner	حرف تعریف
INT	Interjection	حرف صوت
N	Noun	اسم
NUM	Number	عدد
P	Preposition	حرف اضافه
PRO	Pronoun	ضمیر
PUNC	Punctuation	جداکننده
V	Verb	فعل

۳) شناسایی لغت اصلی[1] پرسش: لغت اصلی یک پرسش، بیانگر اطلاعاتی است که کاربر با طرح پرسش به دنبال آن می‌باشد. به عنوان مثال در پرسش " گیاه بومی مناطق کویری چیست؟"، کلمه "گیاه" واژه اصلی پرسش مطرح شده است. تشخیص واژه اصلی در پیاده‌سازی الگوریتم‌های دسته‌بندی متون کاربرد بسیاری دارد [۲۵].

۴) درخت نحوی[2] اجزای جمله: در برخی از زبان‌ها مانند انگلیسی می‌توان با استفاده از دستورات زبان، برای هر یک از جملات یک درخت نحوی رسم نمود و با بررسی ساختار نحوی جمله، نوع لغات تشکیل دهنده آن را تعیین نمود [۲۵]. شکل ۲-۱ نمونه‌ای از درخت تجزیه را نشان می‌دهد.

[1] Headword

[2] Syntactic and semantic parse tree



شکل ۲-۱ درخت تجزیه جمله [۴]

علاوه بر موارد فوق، توجه به روابط معنایی لغات تاثیر فراوانی بر میزان کارایی سیستم‌های بازیابی اطلاعات دارد. از جمله اقداماتی که در تجزیه و تحلیل معنایی جملات صورت می‌پذیرد، می‌توان به موارد زیر اشاره نمود:

۱) تعیین کلمات مترادف و یکسان‌سازی متون: با تعیین و جایگزینی کلمات مترادف، اسناد و پرسش همسان‌سازی شده و در نتیجه دقت پاسخگویی افزایش می‌یابد؛

۲) تعیین اختصارات به کار رفته در متون و نیز اسامی که اختصارات به آنها اشاره می‌کنند؛

۳) تعیین کلمات با معنای عام‌تر[1]: برخی از کلمات معنای عام‌تر و کلی‌تری دارند. به عنوان مثال کلمه "جانوران" نسبت به کلمه "مهره‌داران" معنای عام‌تری دارد، زیرا تمام مهره‌داران جانور هستند در حالی‌که هر جانوری مهره‌دار نیست. از طرف دیگر کلمه "سیب" معنای خاص‌تری[2] نسبت به کلمه "میوه" دارد. استفاده از سطوح عام‌تر به جای سطوح خاص‌تر، نوعی جامعیت را در سطوح بالاتر ایجاد می‌کند که این امر تاثیر بسیاری بر میزان کارایی سیستم‌های بازیابی اطلاعات خواهد داشت [۱۹،۱].

[1] Hypernym
[2] Hyponym

شبکه معنایی، در حقیقت یک پایگاه داده لغوی دسته‌بندی شده می‌باشد. شبکه معنایی اولین بار برای زبان انگلیسی توسعه پیدا کرد که بیش از یکصد هزار مفهوم ارتباط داده شده بوسیله روابط معنایی را شامل می‌شد. هر مفهوم نشان‌دهنده مجموعه انتزاعی از عناصری می‌باشد که بر اساس مشخصه‌های مشترکشان یک گروه را تشکیل می‌دهند. در این لغت‌نامه، برای تشخیص مفاهیم موجود در دنیای واقعی مفاهیم از کل به جزء شکسته شده تا به یک نمونه خاص ختم شود. از این شبکه برای شناسایی کلمات با معنای عام‌تر یا خاص‌تر و نیز کلمات مترادف استفاده می‌شود.

در مجموع تجزیه و تحلیل ساختار گرامری و معنایی جملات در سیستم‌های بازیابی اطلاعات، سبب درک بهتر پرسش کاربر و در نتیجه افزایش امکان تطبیق پرسش با مجموعه اسناد خواهد شد. به عنوان مثال، اگر از الگوریتم ریشه‌یابی در فرایند بازیابی اطلاعات استفاده شود، اسناد بیشتری مرتبط با پرسش مطرح شده شناخته شده و به کاربر ارائه می‌گردند. اما بکارگیری این تجزیه و تحلیل‌ها پیچیدگی محاسباتی را به شدت افزایش می‌دهد و انجام دادن آن بر روی پرسش دریافت شده از کاربر بسیار زمان‌بر و پرهزینه می‌باشد [۴،۱].

۲-۳-۶ الگوریتم TF*IDF

با توجه به اهمیت متفاوت واژه‌ها و اصطلاحات بکار رفته در یک متن از شاخص "وزن اصطلاح" برای نشان دادن اهمیت هر اصطلاح استفاده می‌شود. در این روش به اصطلاحات مهم‌تر وزن بیشتر و به اصطلاحات دارای اهمیت کمتر، وزن کمتری اختصاص داده می‌شود. برای مشخص کردن وزن یک اصطلاح در یک متن، از ضرب تعداد تکرار آن اصطلاح در متن موردنظر (TF) در معکوس تعداد اسناد دارای آن اصطلاح[1] استفاده می‌شود. باید توجه داشت که هرچه تعداد دفعات تکرار یک اصطلاح در یک متن بیشتر و تعداد اسناد دارای آن اصطلاح کمتر باشد، آن اصطلاح دارای اهمیت و وزن بیشتری خواهد بود. فراوانی مدرک معکوس برای واژه i از رابطه زیر محاسبه می‌شود:

[1] Inverse Document Frequency (IDF)

$$IDF_i = log\frac{N}{n_i} \quad (۳-۲)$$

N، تعداد کل اسناد و n_i تعداد اسناد حاوی واژه i است.

برای بازیابی اسنادی که دارای بیشترین احتمال داشتن پاسخ صحیح می‌باشند، می‌توان از روش TF*IDF استفاده نمود تا میزان ارتباط آنها را با پرسش مورد جستجو توسط کاربر تعیین نمود. به این منظور ابتدا با استفاده از روش TF*IDF، وزن هر یک از لغات موجود در پرسش کاربر در هریک از اسناد محاسبه می‌شود. سپس اسناد بر اساس مجموع وزن همه لغات، مرتب شده و اسناد دارای بیشترین امتیاز بازگردانده می‌شوند. باید توجه داشت که در این روش، وزن یا همان اهمیت یک اصطلاح در یک متن با توجه به تعداد تکرار آن اصطلاح در متن مشخص می‌شود و محل بکار رفتن آن در متن تاثیری در وزن کلمه نمی‌گذارد. دلیل استفاده گسترده از این روش نسبت به سایر روش‌ها، سادگی استفاده از آن و نتایج قابل قبول آن است.

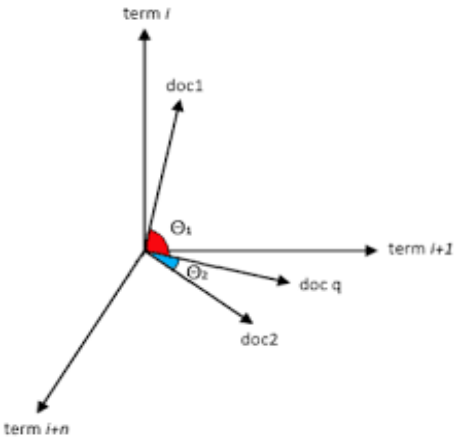
۲-۳-۷ الگوریتم مشابهت کسینوسی[1]

اندازه‌گیری کسینوسی یک روش بسیار معمول اندازه‌گیری تشابه است. در این روش برای محاسبه میزان شباهت بین بردار سند و بردار جستجو، کسینوس زاویه بین دو بردار اندازه‌گیری می‌شود. با اندازه‌گیری تشابه کسینوسی بین هر یک از اسناد با پرسش کاربر و مقایسه مقادیر آنها، مرتبط‌ترین اسناد تعیین و به کاربر ارائه می‌شوند. تشابه کسینوسی برای دو بردار d و q از رابطه زیر محاسبه می‌شود:

$$\mathrm{sim(d,q)} = \frac{d*q}{|d||q|} \quad (۴-۲)$$

در رابطه فوق، d*q حاصلضرب برداری d و q است که با ضرب کردن فرایندهای متناظر در هم، محاسبه می‌شود. شکل ۲-۲ نمایش برداری دو سند و یک پرسش را نشان می‌دهد.

[1] Cosine similarity



شکل ۲-۲ نمایش برداری دو مدرک و یک جستجو

با توجه به این شکل، تشابه میان هر یک از اسناد doc_1 و doc_2 با جستجوی Q، به ترتیب برابر با کسینوس زاویه بین دو بردار یعنی θ_1 و θ_2 است.

$$sim(doc_1, Q) = Cos(\theta_1) \quad (۳-۲)$$
$$sim(doc_2, Q) = Cos(\theta_2)$$

فصل سوم

سیستم‌های پرسش و پاسخ

۳-۱ مقدمه

یکی از روش‌های بازیابی اطلاعات از میان حجم وسیعی از اسناد، استفاده از سیستم‌های پرسش و پاسخ متنی می‌باشد. پرسش و پاسخ، زیرشاخه‌ای از بازیابی اطلاعات است که به ازای پرسش مطرح شده به زبان طبیعی، پاسخ مستقیم را از میان اسناد مرتبط استخراج نموده و در قالب لیست کوتاهی از جملات و یا قطعات کوتاهی از متن به کاربر ارائه می‌نماید. در زمینه بازیابی اطلاعات، تمرکز سیستم‌های پرسش و پاسخ، یافتن روش‌ها و مدل‌هایی است که با انجام تحلیل‌ها، پردازش‌ها و پیش‌پردازش‌های لازم بر روی منابع اطلاعاتی که عمدتا منابع متنی می‌باشند، پاسخی دقیق و مستقیم برای پرسش‌های ارائه شده تولید و استخراج نمایند. در این فصل به بیان انواع سیستم‌های پرسش و پاسخ، معماری و تشریح اجزای تشکیل دهنده، الگوریتم‌های مورد استفاده و روش‌های ارزیابی موجود در این سیستم‌ها پرداخته شده است.

۳-۲ انواع سیستم‌های پرسش و پاسخ

سیستم‌های پرسش و پاسخ می‌توانند از دیدگاه‌های مختلفی دسته‌بندی شوند که در ادامه به توضیح سه نوع دسته‌بندی پرداخته شده است:

۱) انواع سیستم‌های پرسش و پاسخ از دیدگاه حوزه دانش و کاربرد [۲۶]

- سیستم‌های پرسش و پاسخ با دامنه محدود[1]: این سیستم‌ها، سیستم‌هایی هستند که محدوده اطلاعاتی خاصی را شامل می‌شوند. این سیستم‌ها از منابع دانش محدودی استفاده کرده و کاربران مجازند پرسش‌های خود را تنها در آن زمینه خاص مطرح نمایند. با توجه به افزونگی داده در متون، انتخاب الگوریتم مناسب برای این سیستم‌ها جهت انتخاب پاراگراف دربردارنده پاسخ از اهمیت فراوانی برخوردار است. سامانه پرسش و پاسخ پزشکی و نیز سامانه

[1] Restricted domain

پرسش و پاسخ اطلاعات جغرافیایی، از نمونه‌های سیستم‌های پرسش و پاسخ با دامنه محدود می‌باشند.

- سیستم‌های پرسش و پاسخ با دامنه باز[1]: هدف این سیستم‌ها پاسخگویی به پرسش‌های مطرح شده به صورت کلی و بدون وجود پیش فرض در زمینه موضوع سوال و دامنه کاربرد می‌باشد. بدیهی است چنین فرضی به ایجاد یک مدل پرسش و پاسخ عمومی منجر خواهد شد که تنها شرط ضروری آن وجود منابع اطلاعاتی گسترده، مرتبط با انواع پرسش‌های مطرح شده است.

۲) انواع سیستم‌های پرسش و پاسخ از دیدگاه نوع پرسش و پاسخ [۲۸،۲۷]

- سیستم‌های پرسش و پاسخ مبتنی بر حقایق[2]: هدف از طراحی این سیستم‌ها، پاسخگویی کوتاه به پرسش‌هایی است که کاربر در مورد حقایق مطرح می‌نماید. ویژگی اصلی این پرسش‌ها، وجود کلمات پرسشی مانند چه کسی، چه تاریخی و چه مکانی، در سوال کاربر می‌باشد. پاسخ این‌گونه پرسش‌ها حقیقتی کوتاه مانند نام اشخاص، سازمان‌ها، موجودیت‌ها، تاریخ و غیره است که مستقیما در متن ذکر شده و قابل استخراج است.
- سیستم‌های پرسش و پاسخ با قابلیت استنتاج ساده[3]: این سیستم‌ها می‌توانند از نوع ساده‌ای از استدلال جهت پاسخگویی به پرسش‌هایی که پاسخ آن‌ها مستقیما در متن وجود ندارد، استفاده کنند. به عنوان مثال این سیستم‌ها می‌توانند با استفاده از این استدلال ساده به سوالات دارای کلمه "چگونه" پاسخ دهند.
- سیستم‌های پرسش و پاسخ با قابلیت ادغام[4]: این سیستم‌ها می‌توانند پاسخ‌های به دست آمده از چندین سند را با هم ترکیب نموده و آن‌ها را به صورت یک پاسخ به کاربر ارائه نمایند.

[1] Open domain
[2] Factoid QA
[3] Simple reasoning QA
[4] Answer fusion QA

- سیستم‌های پرسش و پاسخ تعاملی[1]: در این سیستم‌ها امکان تعامل دو سویه بین کاربر و سیستم با هدف افزایش دقت پاسخگویی فراهم شده است.
- سیستم‌های پرسش و پاسخ با قابلیت استنتاج پیچیده[2]: در این سیستم‌ها پاسخ پرسش کاربر مستقیما در اسناد ذکر نشده است و سیستم جهت پاسخگویی به سوال کاربر نیازمند استدلال پیچیده‌تری است. به عنوان مثال پاسخ به سوال "چه تیمی در این فصل قهرمان خواهد شد؟" نیازمند تفکر و استدلال پیچیده از سوی سیستم می‌باشد.

۳) انواع سیستم‌های پرسش و پاسخ از دیدگاه دسترسی به اطلاعات

- سیستم‌های مبتنی بر پایگاه دانش: در این سیستم‌ها دانش مورد نیاز جهت پاسخگویی به سوالات کاربر، به شکل پایگاه دانش جمع‌آوری شده است. چنین سامانه‌هایی تنها در مواردی قادر به پاسخگویی هستند که اطلاعات مربوط به دامنه مورد بررسی آن‌ها در قالب یک پایگاه دانش منسجم گردآوری شده باشد. مزیت چنین راهکاری در این است که یک مدل مفهومی برای دامنه کاربرد مورد استفاده در قالب پایگاه دانش وجود دارد. چنین سیستمی قادر است با استفاده از روش‌های پیشرفته‌ای نظیر اثبات قضیه و استدلال به درخواست‌های اطلاعاتی پیچیده پاسخ دهد. در مقابل، چنین سیستمی ما را به مواردی که دانش مورد نیاز از قبل به صورت پایگاه دانش وجود داشته باشد محدود می‌سازد. چالش اصلی این نوع سیستم، تبدیل سوالات مطرح شده به زبان طبیعی، به یک پرس‌وجو می‌باشد. این سیستم‌ها برای پرسش و پاسخ در یک دامنه محدود که در آن امکان مدل‌سازی و ذخیره دانش دامنه مورد بررسی در یک پایگاه دانش وجود دارد، مناسب می‌باشند. BASEBALL و LUNAR دو نمونه از سیستم‌های پرسش و پاسخ مبتنی بر پایگاه دانش هستند.
- سیستم‌های مبتنی بر متن: در اکثر سیستم‌های پرسش و پاسخ، منابع داده موجود متون

[1] Interactive QA

[2] Analogical reasoning QA

ساختارنیافته‌ای مانند روزنامه‌ها، مجلات، مقالات و غیره می‌باشد. در این سیستم‌ها پرسش کاربر با قطعاتی از متن مانند عبارات و جملات مقایسه می‌شود و پاسخ از میان آن‌ها استخراج می‌گردد. در سیستم‌های پرسش و پاسخ مبتنی بر متن، افزونگی داده نقش مهمی در استخراج پاسخ دارد و اطلاعاتی که با ساختار مشابه سوال در متن ظاهر شده باشند دارای احتمال بیشتری برای ارائه به عنوان پاسخ می‌باشند. در این نوع سیستم‌ها، استفاده از روش‌های پردازش زبان طبیعی مانند شناسایی موجودیت‌های با نام و دسته‌بندی مرسوم است. از سوی دیگر افزایش حجم متون، هزینه محاسبات جهت یافتن پاسخ مناسب را افزایش می‌دهد.

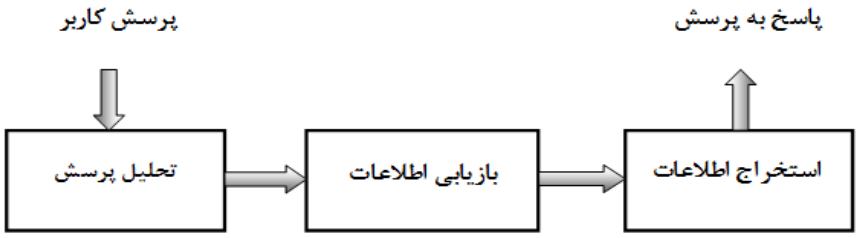
۳-۳ معماری سیستم‌های پرسش و پاسخ

هر چند در طراحی سیستم‌های پرسش و پاسخ تفاوت‌های جزئی وجود دارد، اما اصول اولیه این سیستم‌ها یکسان است. اکثریت سیستم‌های پرسش و پاسخ شامل دو بخش اصلی بازیابی اطلاعات و استخراج اطلاعات می‌باشند [۷].

بخش بازیابی اطلاعات پرسش کاربر را دریافت کرده و با استفاده از تکنیک‌های موجود، اسناد یا قطعه متونی را که احتمال می‌دهد با پرسش مطرح شده ارتباط داشته و یا دربردارنده پاسخ آن باشند بازمی‌گرداند تا دامنه جستجو محدود شود. سپس بخش استخراج اطلاعات، اسنادی را که از مرحله قبل به دست آمده‌اند، پردازش نموده و از بین قطعه متون بررسی شده، محتمل‌ترین پاسخ را استخراج و به عنوان پاسخ نهایی به کاربر ارائه می‌نماید [۷،۵].

در فرایند بازیابی اطلاعات، معمولا کلمات پرسش صرف‌نظر از معنا و گرامر آن برای پرس‌وجو به سیستم بازیابی اطلاعات ارائه می‌شوند تا اسناد و قطعات متنی مرتبط بازیابی شوند. با این وجود سیستم‌های پرسش و پاسخی وجود دارند که در این مرحله از تکنیک‌های پیچیده‌تر پردازش زبان طبیعی استفاده می‌کنند تا با استخراج ویژگی‌های ساختاری، گرامری و معنای کلمات موجود در پرسش نتایج دقیق‌تری را در بازیابی اسناد به دست آورند. برخی از طراحان سیستم، پیش‌پردازش

انجام شده بر روی پرسش مطرح شده توسط کاربر را به عنوان بخش جدیدی در نظر می‌گیرند. این بخش که قبل از بخش بازیابی اطلاعات قرار گرفته است، بخش "تحلیل پرسش" نامیده می‌شود [۲۰،۱۹]. شکل ۳-۱ ساختار سیستم‌های پرسش و پاسخ را نشان می‌دهد.



شکل۳-۱ ساختار سیستم‌های پرسش و پاسخ

۳-۳-۱ بخش اول: تحلیل پرسش

اولین اقدام سیستم پرسش و پاسخ پس از دریافت سوال کاربر، تحلیل پرسش می‌باشد که خود از مراحل مختلفی تشکیل شده است که در ادامه توضیح داده می‌شوند. باید توجه داشت که بعضی از این مراحل را می‌توان بصورت اختیاری حذف نمود [۲۵،۲۴].

- علائمی مانند ویرگول، نقطه، علامت تعجب و غیره از پرسش حذف می‌شوند؛
- کلمات کلیدی پرسش استخراج می‌شوند؛
- کلمات عمومی و فاقد ارزش از مجموعه کلمات حذف می‌شوند؛
- لغات و افعال موجود در پرسش ریشه‌یابی می‌شوند؛
- با استفاده از شبکه معنایی، کلمات هم‌معنی با کلمات موجود در پرسش استخراج می‌شود تا از آن‌ها برای گسترش عبارت پرس‌وجو و افزایش دقت پاسخ استفاده گردد؛
- با استفاده از پایگاه دانش و قوانین از پیش تعریف شده و یا روش‌های مبتنی بر یادگیری ماشین، پرسش مطرح شده دسته‌بندی شده و نوع آن مشخص می‌شود؛
- در حالت‌های پیشرفته‌تر، شناسایی و استخراج روابط معنایی بین عناصر پرسش نیز انجام

می‌شود.

در بین مراحل فوق، تعیین نوع پرسش از اهمیت خاصی برای انتخاب و بازگرداندن پاسخ دقیق برخوردار است. به عنوان مثال، اگر در این مرحله نوع پرسش کمی تشخیص داده شود، سیستم از میان پاسخ‌های محتمل برای یک پرسش، پاسخی را که عدد است بازمی‌گرداند. با توجه به اهمیت این مرحله، چگونگی دسته‌بندی‌بندی پرسش‌ها و الگوریتم‌های بکار رفته برای آن شرح داده می‌شود.

۳-۳-۱-۱ دسته‌بندی پرسش‌ها

در مرحله دسته‌بندی پرسش‌ها، هر پرسش به یک دسته خاص که مشخص‌کننده نوع پاسخ احتمالی می‌باشد نگاشت می‌شود. از آنجا که اگر پرسشی در دسته اشتباه قرار بگیرد توانایی دریافت کردن پاسخ صحیح را از دست می‌دهد، انجام صحیح و دقیق این عملیات از اهمیت زیادی برخوردار است. طبقه‌بندی[1] پرسش‌ها باید بتواند انواع گوناگون پرسش‌ها را شامل شود. یکی از مشهورترین دسته‌بندی‌های پرسش، دسته‌بندی ارائه شده توسط لی و روث[2] در سال ۲۰۰۶ می‌باشد. این محققان الگوریتم طبقه‌بندی خاصی را بر اساس ویژگی‌های ساختاری، معنایی، کلمات مترادف، و فهرست کلمات مربوط به هر نوع پرسش که به صورت دستی برچسب خورده بودند، ارائه نمودند و آن را با استفاده از پایگاه دادگان TREC 10-11 بر روی ۱۰۰۰ پرسش ارزیابی کردند. طبقه‌بندی حاصل، یک طبقه‌بندی سلسله مراتبی ۲ سطحی شامل ۶ کلاس درشت‌دانه[3] و ۵۰ کلاس ریزدانه[4] بود. این محققان به دقت ۹۲,۵٪ در درشت‌دانه‌ها و دقت ۸۵٪ در ریزدانه‌ها دست یافتند [۲۵،۲۱،۴،۱]. جدول

[1] Taxonomy
[2] Li and Roth
[3] Coarse-grained
[4] Fine-grained

۳-۱ دسته‌بندی ارائه شده توسط این محققان را که شامل ۶ دسته درشت‌دانه و ۵۰ دسته ریزدانه می‌باشد، نمایش می‌دهد.

جدول ۳-۱ کلاس‌های درشت‌دانه و ریزدانه [۴]

درشت‌دانه	ریزدانه
ABBREVIATION	Abb – Exp
ENTITY	Animal – Body – Color – Creative – Currency - Dis. Med. – Event – Food – Instrument – Lang – Letter – Other – Plant – Product – Religion – Sport – Substance – Symbol – Technique – Term – Vehicle - Word
DESCRIPTION	Definition – Description – Manner – Reason
HUMAN	Group – Ind – Title – Description
LOCATION	City – Country – Mountain – Other – State
NUMERIC	Code – Count – Date – Distance – Money – Order – Other – Period – Percent – Speed – Temp – Size – Weight

۳-۳-۱-۱-۱ الگوریتم دسته‌بندی پرسش

روش‌های موجود جهت دسته‌بندی پرسش‌ها به دو گروه کلی تقسیم می‌شوند: ۱) روش‌های مبتنی بر قانون ۲) روش‌های مبتنی بر یادگیری ماشین [۱]. در روش‌های مبتنی بر قانون، نوع پرسش با توجه به قوانین از پیش نوشته شده‌ای تعیین می‌شود. این روش، نیازمند قوانین از پیش تعریف شده زیادی است که برای هر مجموعه از سوالات نیز متفاوت می‌باشد و استفاده از آن‌ها در دادگان دیگر، کارایی الگوریتم را کاهش خواهد داد.

در روش‌های مبتنی بر یادگیری ماشین، دسته‌بندی پرسش‌ها در سه مرحله انجام خواهد شد: ۱) استخراج ویژگی ۲) آموزش طبقه‌بند ۳) پیش‌بینی نوع سوالات با استفاده از طبقه‌بند آموزش دیده. مساله مهم در دسته‌بندی سوالات با این روش، استخراج ویژگی‌های مناسب جهت آموزش طبقه‌بند است. ویژگی‌های مورد استفاده جهت تعیین نوع سوال، به سه دسته عمده تقسیم می‌شوند: ۱) ویژگی‌های لغوی ۲) ویژگی‌های نحوی ۳) ویژگی‌های معنایی [۲۵]. از جمله ویژگی‌های لغوی می‌توان به ویژگی‌های N-گرام، کلمات پرسشی، تعداد کلمات موجود در پرسش اشاره نمود. از جمله

ویژگی‌های نحوی می‌توان به ویژگی نوع کلمات بکار رفته در پرسش و نیز تعیین کلمه اصلی پرسش (که دربردارنده موضوع سوال است) اشاره نمود. این ویژگی‌ها عمدتا بر اساس درخت نحوی و بررسی ساختار جمله با استفاده از مجموعه‌ای از قوانین مبتنی بر گرامر استخراج می‌شوند. از جمله ویژگی‌های معنایی می‌توان به ویژگی‌های تعیین کلمات با معنای عام‌تر، و تعیین کلمات هماهنگ اشاره نمود. کلمات هماهنگ به کلماتی گفته می‌شود که یا مانند روزهای هفته از نظر معنایی با هم مرتبط هستند و یا مانند کلمات "دسته"، "گروه" و "طبقه" به شکل‌های مختلفی ظاهر می‌شوند.

برای آموزش طبقه‌بند از مجموعه‌ای از پرسش‌ها که بصورت دستی برچسب خورده‌اند و هر برچسب نوع پاسخ مناسب برای هر پرسش را نشان می‌دهد، استفاده می‌گردد. یادگیرنده‌های قانون و درخت، ماشین بردار پشتیبان[1]، k-نزدیکترین همسایه[2] و طبقه‌بند بیزین[3] مشهورترین الگوریتم‌ها در دسته‌بندی پرسش‌ها می‌باشند [۲۵].

در سال ۲۰۱۳، ین و همکارانش [۲۱] از الگوریتم SVM جهت دسته‌بندی پرسش‌ها استفاده نمودند. این محققان از چهار ویژگی برای دسته‌بندی پرسش‌ها استفاده نمودند که عبارتند از: ۱) کلمات پرسش ۲) ۲-گرام‌ها در پرسش ۳) کلمات دارای معنای عام‌تر که با استفاده از شبکه معنایی استخراج شده بودند ۴) کلمات هماهنگ[4]. البته با توجه به افزایش پیچیدگی محاسبات همزمان با استخراج کلمات هماهنگ و کلمات عام، این کار فقط برای فهرستی از ۵۰ کلمه که دارای بیشترین اهمیت بودند صورت پذیرفت. در ارزیابی خوشه‌های کلمات هماهنگ و عام مربوط به این ۵۰ کلمه، این محققان به دقت ۹۳٪ در درشت‌دانه‌ها و دقت ۸۸,۶٪ در ریزدانه‌ها دست یافتند. در این الگوریتم از دسته‌بندی پیشنهاد شده توسط لی و روث جهت پوشش نوع پرسش‌ها استفاده گردید و به هیچ یک از پرسش‌ها برچسبی غیر از برچسب‌های موجود در این دسته‌بندی داده نشد.

[1] Support Vector Machine (SVM)

[2] K-Nearest Neighbour (KNN)

[3] Naïve bayes

[4] Coordinate

۳-۳-۲ بخش دوم: بازیابی اطلاعات

در بسیاری از سیستم‌های پرسش و پاسخ، منبع اطلاعات شامل مجموعه بزرگی از اسناد متنی است که جستجو و کاوش تمام این مجموعه جهت یافتن پاسخ مناسب دشوار و زمان‌بر است. به همین دلیل در بسیاری از سیستم‌های پرسش و پاسخ به ویژه سیستم‌های پرسش و پاسخ با دامنه باز مولفه‌ای جهت بازیابی اطلاعات وجود دارد [۲۴] که فیلتر کردن اسناد موجود را جهت کوچک نمودن مجموعه اسناد دارای پاسخ احتمالی انجام می‌دهد. پس از ورود مخزنی از اسناد و اطلاعات متنی به همراه پرسش کاربر به این مولفه، تمامی اسناد از طریق یک تابع و بر اساس میزان ارتباطشان با پرسش کاربر، امتیازدهی و مرتب می‌شوند. بدیهی است صحت پاسخ استخراج شده در مراحل بعدی، به صحت اسناد بازیابی شده در مولفه بستگی دارد که این امر نشان‌دهنده اهمیت بسیار زیاد مولفه بازیابی اطلاعات در سیستم‌های پرسش و پاسخ است. در این بازیابی ممکن است از روش‌های مختلفی از جمله روش TF*IDF، مشابهت کسینوسی و مدل احتمالاتی استفاده شود [۷].

۳-۳-۲-۱ بازیابی پاراگراف

یکی از چالش‌های بازیابی اطلاعات حجم متون بازیابی شده است، زیرا برای تشخیص پاسخ مرتبط با یک پرسش، بخش کوچکی از سند هم کافی است و بازیابی محتوای کل سند زمان پاسخگویی به پرسش را افزایش می‌دهد و حتی به دلیل وجود نویز فراوان، ممکن است سیستم را از پاسخ اصلی دور نماید [۲۴]. بنابراین باید از مولفه‌ای برای بازیابی اطلاعات استفاده شود که طول محدودتری از اسناد را ارائه دهد که احتمال وجود پاسخ در آن بسیار زیاد است.

در روش‌های اولیه بازیابی پاراگراف، تعداد لغات مشترک بین پاراگرف و سوال کاربر تعیین شده و بر اساس آن پاراگراف‌های مناسب احتمالی بازگردانده می‌شوند. این روش می‌تواند منجر به بازگرداندن پاراگراف‌هایی شود که با سوال مرتبط بوده اما شامل جواب صحیح نیستند. برای رفع این مشکل، به تدریج ویژگی‌های دیگری نیز در بازیابی پاراگراف در نظر گرفته شدند. مساله مهم در بازیابی پاراگراف،

انتخاب ویژگی‌های مناسب برای امتیازدهی پاراگراف‌ها است. از ویژگی‌های مورد استفاده در بازیابی پاراگراف می‌توان به موارد زیر اشاره نمود [۷]:

۱) تعداد موجودیت‌های یافت شده در پاراگراف که با نوع پرسش مطابقت دارند

۲) تعداد کلمات مشترک با پرس وجوی کاربر

۳) تعداد N-گرام‌های مشترک با پرسش کاربر

۴) فاصله بین کلمات کلیدی مورد جستجو در پاراگراف

در پاسخگویی به سوالات پیچیده[1] عمدتا پاسخ دقیق استخراج نمی‌شود و اکثر سیستم‌های پرسش و پاسخ، پاراگراف مرتبط را بازمی‌گردانند. بعلاوه، بعضی از سیستم‌های پرسش و پاسخ، در پاسخ به سوالات حقیقی[2] به جای بازگرداندن کلمه و یا عبارات دقیق، بخش کوتاهی از اطلاعات مانند پاراگراف و یا جمله را بازمی‌گردانند و تنها بخشی از آن را که به احتمال زیاد پاسخ دقیق پرسش می‌باشد برجسته می‌نمایند [۲۱]. از آنجا که استخراج پاسخ صحیح در گام بعدی سیستم پرسش و پاسخ، وابسته به صحت پاراگراف بازیابی شده است، این مولفه از اهمیت زیادی برخوردار است. الگوریتم‌های مختلفی جهت بازیابی پاراگراف وجود دارند که از آن جمله می‌توان به روش‌های TF*IDF، مشابهت کسینوسی و مدل احتمالاتی اشاره نمود. در سیستم طراحی شده، از الگوریتم JIRS مبتنی بر تکنیک‌های آماری جهت بازگرداندن پاراگراف استفاده شده که توضیحات کامل آن در بخش ۵-۲-۵ آمده است.

۳-۳-۳ بخش سوم: استخراج اطلاعات

بخش استخراج اطلاعات که آخرین بخش سیستم پرسش و پاسخ را تشکیل می‌دهد، قطعه‌های متنی بازگردانده شده از مرحله قبل را پردازش نموده و عباراتی که احتمال می‌رود دارای پاسخ دقیق

[1] Complex questions

[2] Factoid questions

باشند استخراج می‌نماید. سپس عبارتی که بیشترین احتمال دارا بودن پاسخ را دارد از بین عبارات مذکور انتخاب شده و به عنوان پاسخ نهایی سیستم به کاربر ارائه می‌شود.

جهت استخراج پاسخ، بخش استخراج اطلاعات در سیستم غالبا نیازمند بکارگیری دادگان آموزشی و همچنین استفاده از تکنیک‌های یادگیری ماشین می‌باشد. در این روش‌ها، الگوریتم‌های یادگیری ماشین بر روی مجموعه حاشیه‌نویسی شده از پرسش و پاسخ‌ها که یک مجموعه آموزشی محسوب می‌شود، آموزش می‌بینند و پس از آن قادر به استخراج اطلاعات بر اساس ویژگی‌های آموزش دیده می‌باشند. ویژگی‌هایی که در روش یادگیری ماشین بیشتر مورد توجه هستند عبارتند از [۲۱]:

- تطابق نوع پاسخ
- تطابق الگو: با توجه به تطابق عبارت منظم حاشیه‌نویسی شده و پاسخ‌های کاندید
- تعداد کلمات کلیدی مشترک با سوال
- تعداد کلمه موجود بین پاسخ‌های کاندید و کلمات کلیدی سوال
- کلماتی که در پاسخ کاندید وجود دارند، اما در سوال وجود ندارند
- نشانه‌گذاری‌های موجود بلافاصله پس از پاسخ کاندید
- N–گرام‌های شامل کلمات پرسش موجود در پاسخ کاندید

۳-۴ معیارهای ارزیابی سیستم‌های پرسش و پاسخ

امروزه ارزیابی سیستم‌های پرسش و پاسخ برای تعیین میزان موفقیت و کارایی آن‌ها به یکی از حوزه‌های پژوهشی مهم در علم رایانه تبدیل شده است. سیستم‌های پرسش و پاسخ با توجه به مولفه استخراج اطلاعات و رویکرد اتخاذ شده برای انتخاب پاسخ، به دسته‌های مختلفی تقسیم می‌شوند که مهم‌ترین آن‌ها عبارتند از:

۱) سیستم‌هایی که به دلایلی مانند عدم اطمینان کافی به بعضی از سوالات پاسخی نمی‌دهند؛

۲) سیستم‌هایی که برای هر پرسش پاسخی برمی‌گردانند.

پاسخ سیستم می‌تواند به یکی از اشکال زیر باشد:

۱) برخی از سیستم‌ها در پاسخ به یک پرسش، فهرستی از پاسخ‌ها را که با توجه به امتیاز کسب شده رتبه‌بندی شده‌اند بازمی‌گردانند.

۲) برخی از سیستم‌ها برای هر پرسش کاربر تنها یک پاسخ بازمی‌گردانند.

با در نظر گرفتن دسته‌بندی‌های فوق، معیارهای متعددی جهت ارزیابی نظام پرسش و پاسخ مطرح شده است که برخی از آن‌ها بسیار رایج و مرسوم هستند. قبل از ارائه این معیارها تعدادی از متغیرهای بکار رفته در آن‌ها معرفی می‌شود:

- Q: به مجموعه پرسش‌هایی که به مولفه بازیابی اطلاعات در سیستم پرسش و پاسخ داده می‌شود اشاره می‌کند.
- q: نشان‌دهنده هر یک از پرسش‌های موجود در مجموعه پرسش‌ها می‌باشد.
- D: نشان‌دهنده مجموعه متون موجود در منابع اطلاعات سیستم می‌باشد.
- $A_{(q,D)}$: به مجموعه قطعات متنی در منبع اطلاعات که پاسخ‌های صحیحی برای پرسش q می‌باشند اشاره می‌کند.
- S: زیرمجموعه‌ای از اسناد موجود در منبع اطلاعاتی D است که در ارتباط با پرسش کاربر ارزیابی می‌شوند.
- $R_{(q,D,n)}{}^{S}$: هنگامی که پرسش q به مولفه بازیابی اطلاعات وارد می‌شود، مجموعه اسناد S از منبع اطلاعاتی D بازیابی می‌شوند. در این مجموعه، n سند یا قطعه متن بازیابی‌شده که دارای بیشترین امتیاز هستند، $R_{(q,D,n)}{}^{S}$ نامیده می‌شوند.

در ادامه تعدادی از معیارهای ارزیابی که به صورت گسترده استفاده می‌شوند، آورده شده است:

۱) میانگین رتبه متقابل[1]: هنگامی که پرسش q_1 تحویل مولفه بازیابی اطلاعات می‌شود، این مولفه زیرمجموعه‌ای از پاسخ‌ها را که به نام S_{q_1} شناخته می‌شود، بازیابی می‌کند. چنانچه در

[1] Mean Reciprocal Rank (MRR)

مجموعه پاسخ‌های بازیابی شده (S_{q_1})، پاسخ صحیح، پاسخ j ام باشد، امتیاز پرسش با استفاده از فرمول $Rank_{q_1} = \frac{1}{\mathrm{j}}$ تعیین می‌شود. هنگامی که این محاسبه برای تمام پرسش‌ها در مجموعه Q انجام شود و میانگین امتیازات کسب شده برای تک‌تک پرسش‌ها محاسبه گردد، مقدار جدیدی که به نام MRR شناخته می‌شود، به دست می‌آید. این مقدار برابر با میانگین معکوس رتبه‌های پاسخ‌های ارائه شده صحیح برای پرسش‌های مجموعه Q می‌باشد [۲۳،۵،۱].

$$\mathrm{MRR} = \frac{1}{\mathrm{Q}} \sum_{i=1}^{Q} Rank_{q_i} \qquad (۱-۳)$$

عملکرد سیستم در ارائه پاسخ‌های صحیح با کمیت MRR ارتباط مستقیم دارد و با افزایش کمیت MRR، عملکرد سیستم بهتر خواهد بود، زیرا بالاتر بودن میزان MRR نشان‌دهنده بالاتر بودن امتیاز محاسبه شده برای یکایک پرسش‌های Q و بنابراین کوچکتر بودن مخرج j در فرمول ($Rank_{q_{\mathrm{i}}} = \frac{1}{\mathrm{j}}$) می‌باشد. کمتر بودن مخرج در این فرمول بدین معناست که پاسخ صحیح به ابتدای فهرست پاسخ‌های بازیابی شده نزدیکتر بوده و بنابراین سرعت ارائه پاسخ صحیح به کاربر افزایش می‌یابد. یکی از اشکالات این معیار این است که با استفاده از آن امکان ارزیابی پرسش‌هایی که همه پاسخ‌های پیشنهاد شده برای آن‌ها نادرست بوده‌اند، وجود ندارد و بنابراین کمیت MRR به میزان بازخوانی سیستم توجه نمی‌کند [۱]. اشکال دیگر این است که نمی‌توان از این معیار در ارزیابی سیستم‌های پرسش و پاسخی که برای هر پرسش فقط یک پاسخ ارائه می‌کنند استفاده نمود، زیرا در این‌گونه سیستم‌ها فهرستی از پاسخ‌ها که بر اساس امتیاز مرتب شده‌اند، به کاربر ارائه نمی‌شود.

۲) بازخوانی[1] و دقت[2]: با در نظر گرفتن اشکالات ذکر شده، مشخص می‌شود معیار MRR معیار مناسبی برای ارزیابی پرسش‌هایی که فاقد پاسخ صحیح در مجموعه بازیابی شده هستند، نمی‌باشد.

[1] Recall

[2] Precision

به همین دلیل، دو معیار بازخوانی و دقت جهت ارزیابی کارایی[1] سیستم مورد استفاده قرار می‌گیرند. بازخوانی یعنی از میان کل پاسخ‌های صحیح موجود در مجموعه اسناد، چند درصد از آن‌ها یافت شده‌اند و دقت یعنی در میان پاسخ‌های یافت‌شده، چند درصد صحیح می‌باشند. دقت و بازیابی با استفاده از روابط زیر قابل اندازه‌گیری هستند [۳۱،۳۰،۲۹]:

$$\text{Precision} = \frac{\text{تعداد پاسخ‌های بازیابی شده مرتبط}}{\text{تعداد پاسخ‌های بازیابی شده}} \quad (۲-۳)$$

$$\text{Recall} = \frac{\text{تعداد پاسخ‌های مرتبط بازیابی شده}}{\text{تعداد پاسخ‌های مرتبط}} \quad (۳-۳)$$

جدول ۳-۲ انواع پاسخ‌ها در سیستم پرسش و پاسخ را نمایش می‌دهد.

جدول ۳-۲ انواع پاسخ‌ها در سیستم پرسش و پاسخ [۳۱]

	مرتبط	نامرتبط
بازیابی شده	مثبت واقعی TP	مثبت کاذب FP
بازیابی نشده	منفی کاذب FN	منفی واقعی TN

با توجه به جدول فوق، تعریف دیگری از دقت و بازخوانی به شکل زیر می‌باشد.

$$\text{Precision} = \frac{TP}{\text{TP+FP}} \quad (۴-۳)$$

$$\text{Recall} = \frac{TP}{\text{TP+FN}} \quad (۵-۳)$$

مقدار این دو پارامتر غالبا نسبت معکوسی با هم دارند و افزایش مقدار یکی باعث کاهش و افت دیگری می‌شود؛ از این‌رو بایستی مصالحه‌ای بین این دو مقدار انجام شود.

۳) معیار اف[2]: معمولا بین دقت و توانایی بازخوانی سیستم رابطه معکوس وجود دارد و این در حالیست که برای بسیاری از کاربران دقت سیستم ممکن است نسبت به توانایی بازخوانی آن، به میزان B برابر ارزش بیشتری داشته باشد. با توجه به این مسئله چنانچه برای معیار دقت وزن یا

[1] Performance

[2] F-measure

ارزش B قائل شویم، معیار اف که ترکیبی از معیارهای دقت و بازخوانی و در حقیقت میانگین هارمونیک دو پارامتر فوق است، به صورت زیر محاسبه می‌شود [۳۰،۵،۱]:

$$\text{F-measure (B)} = \frac{(B^2+1)* \text{Precision.Recall}}{B^2*\text{Precision+Recall}} \qquad (۶-۳)$$

با استفاده از این معیار می‌توان دقت و بازخوانی را در مقابل هم تعدیل کرد [۳۲]. این معیار به ویژه زمانی که پاسخ‌های مرتبط در نظام بازیابی اطلاعات اندک هستند، معمولا به معیار دقت ترجیح داده می‌شود [۳۳] و از معیارهای معمول ارزیابی بخصوص در مواقع کار با مجموعه‌های غیرمتعادل است [۳۴،۳۰] تعیین حداکثر مقدار برای F به منظور یافتن بهترین ترکیب ممکن بین جامعیت و مانعیت می‌باشد [۳۳،۳۱].

۴) دقت در n نتیجه اول (p@n): در تعیین معیار دقت، تمامی پاسخ‌های بازگردانده شده توسط سیستم مورد توجه قرار می‌گیرند و بنابراین این معیار اطلاعات مناسبی درباره تعداد پاسخ‌های صحیح موجود در n پاسخ ابتدای فهرست پیشنهادی پاسخ‌ها ارائه نمی‌کند. به عنوان مثال، ممکن است در فهرست پاسخ‌های پیشنهادی توسط یک سیستم پاسخ‌های صحیح در رتبه‌های بالاتر و در ابتدای فهرست ارائه شده باشند و در سیستم دیگر همین تعداد پاسخ صحیح در رتبه‌های پایین و در اواسط یا اواخر فهرست پاسخ‌های صحیح قرار داشته باشند. در این حالت معیار دقت برای دو سیستم ارزش یکسانی در نظر می‌گیرد.

برای تشخیص تفاوت این دو سیستم از نظر دقت پاسخ‌های ارائه شده باید از معیار دیگری که p@n نامیده می‌شود استفاده نمود [۵]. برای ارزیابی توانایی بازخوانی این دو سیستم نیز تنها n نتیجه ابتدای فهرست بازیابی شده که دارای بیشترین امتیاز هستند، در نظر گرفته می‌شوند.

$$precision_{(\text{Q,D,n})}{}^{\text{S}} = \frac{\sum_{q\in Q} \frac{\left|R_{(\text{q,D,n})}{}^{S} \cap A_{(\text{q,D})}\right|}{\left|R_{(\text{q,D,n})}{}^{S}\right|}}{|Q|} \qquad (۷-۳)$$

$$Recall_{(\text{Q,D,n})}{}^{\text{S}} = \frac{\sum_{q\in Q} \frac{\left|R_{(\text{q,D,n})}{}^{S} \cap A_{(\text{q,D})}\right|}{\left|A_{(\text{q,D})}\right|}}{|Q|} \qquad (۸-۳)$$

۵) صحت[1]: نسبت پرسش‌هایی که سیستم پاسخ صحیح به آن‌ها داده است به کل پرسش‌های مطرح شده را معیار صحت نامند و از رابطه زیر قابل محاسبه است [۱]:

$$\text{Accuracy} = \frac{\text{تعداد پرسشهایی که سیستم پاسخ صحیح داده است}}{\text{تعداد پرسشهای مطرح شده}} \quad (۹-۳)$$

۶) سنجش اطمینان[2]: این پارامتر در ارزیابی سیستم‌های پرسش و پاسخی کاربرد دارد که لیستی از پاسخ‌ها را بر اساس میزان اطمینان به کاربر ارائه می‌نمایند [۳۵،۱]:

$$\text{CWS} = \frac{1}{\text{n}} \sum_{i=1}^{N} \frac{number\ of\ correct\ answers\ in\ first\ i\ Rank}{\text{i}} \quad (۱۰-۳)$$

در فرمول فوق n تعداد پرسش، N تعداد پاسخ‌های بازگردانده شده برای هر پرسش و i موقعیتی در لیست مرتب شده پاسخ‌ها است. صورت کسر داخل سیگما، معادل تعداد پرسش‌هایی است که کاربر مطرح نموده است و پاسخ سیستم به آن در موقعیت i ام از فهرست پاسخ‌های مربوطه، صحیح می‌باشد.

۷) C@1: این پارامتر در ارزیابی سیستم‌های پرسش و پاسخی کاربرد دارد که سیستم می‌تواند در صورت عدم اطمینان، به پرسش کاربر پاسخ ندهد. این رویکرد در سیستم‌هایی مانند سیستم تشخیص پزشکی بسیار مناسب است، زیرا عواقب بازنگرداندن پاسخ به مراتب بهتر از بازگرداندن پاسخ اشتباه می‌باشد و این امر افزایش دقت را با ثابت نگه داشتن میزان بازخوانی سیستم در پی خواهد داشت [۳۵].

$$\text{C@1} = \frac{1}{\text{n}}(n_{\text{R}} + n_{\text{U}} \frac{n_{\text{R}}}{\text{n}}) \quad (۱۱-۳)$$

در فرمول فوق، n_{R} تعداد پرسش‌هایی است که سیستم پاسخ صحیحی به آن‌ها داده است، n_{U} تعداد پرسش‌هایی است که سیستم پاسخی به آن‌ها نداده است و n تعداد کل پرسش‌های مطرح شده است.

[1] Accuracy

[2] Confidence Weighted Score (CWS)

۸) پوشش[1]: این احتمال وجود دارد که دو سیستم پرسش و پاسخ به تعداد یکسانی از سوالات، تعداد یکسانی از پاسخ‌های صحیح را ارائه نمایند و در نتیجه از دقت یکسانی برخوردار باشند، اما تعداد سوالات برخوردار از پاسخ صحیح که می‌توان از آن به توانایی سیستم در پوشش دادن به سوالات با پاسخ صحیح تعبیر نمود در دو سیستم متفاوت باشد. به عنوان مثال ممکن است تعداد ۱۰۰ سوال یکسان از دو سیستم پرسش و پاسخ پرسیده شود و سیستم‌ها نیز برای هر کدام از پرسش‌ها تعدادی قطعه متن را که احتمالا دربردارنده پاسخ هستند بازگردانند. بررسی ۱۰۰ پاسخ ابتدای فهرست بازیابی شده، ممکن است یکی از دو حالت زیر را نشان دهد:

- سیستم S1 توانسته است فقط برای یکی از پرسش‌ها، ۱۰۰ پاسخ صحیح بازگرداند، اما برای سایر سوالات هیچ پاسخ صحیحی ارائه نکرده است.
- سیستم S2 توانسته است برای هر یک از ۱۰۰ سوال پرسیده شده، فقط یک جواب صحیح بازگرداند.

بدیهی است از نظر توانایی بازیابی اطلاعات، سیستم S2 نسبت به سیستم S1 کارایی بهتری دارد، زیرا توانسته است برای تمام پرسش‌ها، حداقل یک پاسخ صحیح ارائه نماید، در حالی‌که سیستم S1 فقط توانسته است به یکی از پرسش‌ها پاسخ صحیح دهد.

$$p_{(\mathrm{q,D,n})}{}^{S_1} = \frac{\sum_{q \in Q} \frac{|R_{(\mathrm{q,D,n})}{}^{S} \cap A_{(\mathrm{q,D})}|}{|R_{(\mathrm{q,D,n})}{}^{S}|}}{|Q|} = \frac{\frac{100}{100} + \frac{0}{100} + \cdots + \frac{0}{100}}{100} = 0.01$$

$$p_{(\mathrm{q,D,n})}{}^{S_2} = \frac{\sum_{q \in Q} \frac{|R_{(\mathrm{q,D,n})}{}^{S} \cap A_{(\mathrm{q,D})}|}{|R_{(\mathrm{q,D,n})}{}^{S}|}}{|Q|} = \frac{\frac{1}{100} + \frac{1}{100} + \cdots + \frac{1}{100}}{100} = 0.01$$

با توجه به توضیحات بیان شده، معیار دیگری مورد نیاز است که بتواند تفاوت دو سیستم ذکر شده را به خوبی نشان دهد که از آن به معیار پوشش تعبیر می‌شود [۴]. به عبارت دیگر معیار

[1] Coverage

پوشش کسری از پرسش‌ها است که در n پاسخ برتر بازیابی شده، حداقل یک پاسخ صحیح داشته باشند [۲۴،۷].

$$Coverage_{(\mathrm{Q,D,n})}{}^{S} = \frac{|q \in Q|\{R_{(\mathrm{q,D,n})}{}^{S} \cap A_{(\mathrm{q,D})} \neq \emptyset\}|}{|Q|} \qquad (۱۲-۳)$$

۹) افزونگی[1]: این معیار نشان‌دهنده میانگین تعداد پاسخ صحیح در مجموعه پاسخ‌های بازگردانده شده برای هر پرسش است. به بیان دیگر این معیار احتمال اینکه مولفه استخراج اطلاعات بتواند برای پرسش مطرح شده، حداقل یک پاسخ صحیح بیابد را محاسبه می‌کند [۷].

$$Redundancy_{(\mathrm{q,D,n})}{}^{S} = \frac{\sum_{q \in Q} |R_{(\mathrm{q,D,n})}{}^{S} \cap A_{(\mathrm{q,D})}|}{|Q|} \qquad (۱۳-۳)$$

ذکر چند نکته در ارزیابی سیستم‌های پرسش و پاسخ قابل توجه است:

۱) برخلاف سیستم‌های پرسش و پاسخی که به سوال کاربر به صورت یک عبارت اسمی کوتاه پاسخ می‌دهند، در سیستم‌های پرسش و پاسخ پیچیده که پاسخ به شکل یک قطعه متن ارائه می‌شود، ارزیابی سیستم غالبا به شکل خودکار امکان‌پذیر نیست و از داور انسانی جهت محاسبه معیارهای فوق استفاده می‌شود.

۲) با توجه به ماهیت پاسخگویی سیستم‌های پرسش و پاسخ مورد بررسی، از معیارهای متفاوتی جهت ارزیابی آن‌ها استفاده می‌شود. به عنوان مثال استاندارد TREC از معیارهای دقت و بازخوانی، استاندارد CLEF از C@1 و سنجش اطمینان و استاندارد INEX از معیار پوشش و صحت جهت ارزیابی سیستم‌های پرسش و پاسخ استفاده می‌کنند.

۳) در سیستم‌هایی که برای هر پرسش مطرح شده فقط یک پاسخ در متن وجود داشته اما سیستم می‌تواند به دلایلی مانند عدم اطمینان کافی به بعضی از سوالات پاسخی ندهد، تعداد پاسخ‌های مرتبط معادل تعداد پرسش‌ها است و $|\{Relevant\}| = |\{all\}|$، در نتیجه [۳۶]:

[1] Redandancy

$$\text{Accuracy=Recall=}\frac{|\{Retrieved\}\cap\{Relevant\}|}{|\{all\}|}=\frac{\text{TP}}{\text{TP+FN}}$$

۴) در سیستم‌هایی که برای هر پرسش مطرح شده فقط یک پاسخ در متن وجود دارد و سیستم برای هر پرسش پاسخی بازمی‌گرداند، تعداد پاسخ‌های بازگردانده شده معادل تعداد پرسش‌های مطرح شده است و $|\{all\}| = |\{Relevant\}| = |\{Retrieved\}|$، در نتیجه [۳۶]:

$$\text{Accuracy= Precision=Recall=F1=}\frac{|\{Retrieved\}\cap\{Relevant\}|}{|\{all\}|}=\frac{\text{TP}}{\text{TP+FP}}$$

فصل چهارم

تعامل در سیستم‌های پرسش‌وپاسخ

۴-۱ مقدمه

توسعه سیستم‌های رایانه‌ای و گسترش استفاده از فناوری اطلاعات در زندگی روزمره باعث شده است تا دستیابی سریع به اطلاعات از اهمیت بالایی برخوردار باشد. با افزایش حجم اطلاعات، کنترل و مدیریت آن مشکل‌تر می‌شود و بنابراین تولید و وجود اطلاعات به تنهایی کافی نیست؛ بلکه باید ابزارهایی برای استفاده از این اطلاعات فراهم شود. یکی از روش‌های بازیابی اطلاعات، استفاده از سیستم‌های پرسش و پاسخ متنی است. انفجار اطلاعات در دنیای امروز، باعث شده است تا جامعه پژوهشگران در این حوزه همواره به دنبال ارائه راهکارهایی جهت بهبود نتایج باشند. یکی از چالش‌های موجود در سیستم‌های پرسش و پاسخ، فقدان تعامل دو سویه بین سیستم و کاربر است. افزودن سطح تعامل به سیستم‌های پرسش و پاسخ با دو هدف صورت می‌گیرد: ۱) اگر پرسش کاربر دارای ابهام باشد، سیستم مکالمه‌ای را با کاربر در جهت رفع ابهام و درک بهتر پرسش آغاز می‌نماید. ۲) چنانچه پاسخ سیستم دلخواه کاربر نباشد و یا کاربر نیاز به اطلاعات بیشتری داشته باشد، کاربر مکالمه‌ای را با سیستم آغاز می‌نماید تا پاسخ دلخواه خود را دریافت کند. در این بخش پس از بیان جزئی‌تر مفاهیم مورد نیاز، روش‌ها و چالش‌های موجود در برقراری تعامل بین کاربر و سیستم‌های پرسش و پاسخ تشریح شده است.

۴-۲ مفهوم تعامل

گفتگوی متقابل بین دو یا چند موجودیت را تعامل می‌گویند. در تعامل گوینده مطلبی می‌گوید و مطلب توسط شنونده تفسیر شده و تصمیم می‌گیرد که چه پاسخی اتخاذ نماید. در نهایت پاسخی به گوینده بازگردانده می‌شود و این چرخه گوش دادن، درک کردن، تصمیم‌گیری و پاسخ تکرار می‌شود. هر چند در مکالمه بین انسان‌ها محدودیتی وجود ندارد، در پیاده‌سازی سیستم پرسش و پاسخ مبتنی بر تعامل محدودیت‌های زیر در نظر گرفته می‌شود [۳۷]:

- دقیقا دو شرکت‌کننده با یکدیگر صحبت می‌کنند؛
- هر یک از شرکت‌کنندگان به نوبت صحبت می‌کند؛
- شرکت‌کنندگان در زمینه مشترکی با یکدیگر صحبت می‌کنند.

۳-۴ انواع تعامل

تعامل بین کاربر و سیستم می‌تواند به یکی از سه شکل زیر باشد [۳۷]:

۱) ابتکار عمل دست سیستم است[1] و سیستم هدایت گفتگو را برعهده دارد. بعنوان مثال سیستم از کاربر می‌خواهد از یک فهرست مشخص، موردی را انتخاب کند. در این نوع مکالمات، سیستم در هر مرحله سوالاتی از کاربر می‌پرسد و کاربر پاسخ می‌دهد. در نتیجه کاربر آزادی کمتری در مکالمه داشته و تنها باید با عبارات کوتاه به سوالات سیستم پاسخ بدهد. اکثر سیستم‌های تجاری از این نوع هستند.

۲) ابتکار عمل دست کاربر است[2] و کاربر هدایت گفتگو را برعهده دارد. سیستم باید بگونه‌ای طراحی شود که توانایی پاسخگویی به سوالات کاربر را داشته باشد. در این سیستم‌ها به علت آزادی بیشتر کاربر در هدایت مکالمه، امکان نامفهوم بودن سوال کاربر برای سیستم وجود دارد. در این موارد سیستم از روش "تایید اطلاعات" برای اطمینان یافتن از صحت اطلاعات ورودی استفاده می‌کند.

۳) کاربر و سیستم هر دو دارای ابتکار عمل هستند[3]. در این حالت با وجود اینکه سیستم هدایت گفتگو را در دست دارد، کاربر می‌تواند در هر زمان مسیر مکالمه را تغییر دهد و سیستم باید به سوالات کاربر پاسخ دهد. مکالمات در این نوع سیستم‌ها، شباهت بیشتری به مکالمات انسان با انسان دارند و امروزه بسیار مورد توجه هستند.

---

[1] System initiative
[2] User initiative
[3] Mixed initiative

۴-۴ هدف از تعامل در سیستم‌های پرسش و پاسخ

تعامل بین کاربر و سیستم پرسش و پاسخ در زمینه طرح پرسش و پاسخگویی به آن بوده و اهداف زیر را برآورده می‌نماید [۳۸]:

- کاربر می‌تواند به جای طرح مجموعه‌ای از چندین سوال مرتبط در قالب یک پرسش پیچیده، آن را به سوالات ساده‌تر و کوتاهتر تجزیه نماید که این امر، مکالمه طبیعی‌تر و پاسخ دقیق‌تر را در پی خواهد داشت.
- چنانچه پاسخ سیستم دلخواه کاربر نباشد و یا کاربر نیاز به اطلاعات بیشتری داشته باشد، مکالمه‌ای را با سیستم آغاز می‌نماید تا پاسخ دلخواه خود را دریافت کند.
- امکان طرح مجموعه‌ای از سوالات مرتبط با هم[1] از سوی کاربر وجود دارد. به عنوان مثال کاربر می‌تواند پرسشی را مطرح کند که از طریق ضمایر به پرسش قبل ارجاع داده شده است.
- اگر پرسش کاربر دارای ابهام باشد، سیستم مکالمه‌ای را با کاربر در جهت رفع ابهام و درک بهتر پرسش آغاز می‌نماید.
- سیستم می‌تواند پس از ارائه پاسخ به یک پرسش، مجموعه‌ای از سوالات مرتبط با آن موضوع را به کاربر پیشنهاد نماید که انسجام گفتگو و افزایش کارایی را در پی خواهد داشت.

۴-۴-۱ ابهام درسیستم‌های پرسش و پاسخ

با توجه به ابهامات موجود در زبان طبیعی، رفع ابهام در تعامل انسان و ماشین اهمیت بسزایی دارد و سبب جلوگیری از شکست در مکالمات می‌شود. ابهام به حالتی گفته می‌شود که یکی از شرکت‌کنندگان قادر به فهم مطلب دیگری نباشد. مثال زیر را در نظر بگیرید:

– کاربر: وزیر امورخارجه ایران در چه زمانی با همتای روسی خود ملاقات کرد؟

[1] Follow up question

– سیستم: ۶ فروردین ۱۳۹۴

– کاربر: در این نشست او چه مطالبی را عنوان کرد؟

در مثال فوق مرجع ضمیر"او" مبهم است و سیستم جهت رفع ابهام، پرسش زیر را از کاربر مطرح می‌نماید.

– سیستم: منظور شما از ضمیر"او"، "وزیر امورخارجه ایران" است یا "وزیر امورخارجه روسیه" ؟

برخی از نشانه‌های وجود ابهام عبارتند از:

- عدم یافتن مرجع مناسب ضمایر توسط سیستم
- وجود کلمات ناآشنا مانند کلمات اختصار
- وجود تناقضات معنایی در پرسش

۲-۴-۴ سوالات مرتبط

چنانچه سیستم پرسش و پاسخ تاریخچه گفتگوها را نگهداری کند، کاربر می‌تواند پرسشی را مطرح کند که به مجموعه سوالات قبلی وی مرتبط باشد. پاسخگویی به این نوع پرسش‌ها عمدتا نیازمند برقراری تعامل بین کاربر و سیستم است. سوالات مرتبط به دو دسته کلی تقسیم می‌شوند:

۱) سوالات حذفی[1]: در این سوالات بخشی از جمله مانند فعل، با توجه به جمله قبلی حذف شده است. مثال زیر را در نظر بگیرید:

– کاربر: شکسپیر در چه سالی متولد شد؟

– سیستم: در سال ۱۵۶۴

– کاربر: کجا؟

۲) سوالات دارای ضمیر[2]: سوالاتی که با استفاده از ضمایر به بخشی از سوال قبل ارجاع داده

[1] Elliptic question

[2] Anaphora

می‌شوند. مثال زیر را در نظر بگیرید:

- کاربر: شکسپیر در چه سالی متولد شد؟
- سیستم: در سال ۱۵۶۴
- کاربر: او با چه کسی ازدواج کرد؟

۴-۵ نمودار جریان در سیستم‌های پرسش و پاسخ تعاملی

- کاربر پرسشی را مطرح می‌نماید؛
- پرسش ورودی بررسی می‌شود که آیا پاسخگویی به آن نیاز به تعامل با کاربر دارد یا خیر؛
- در صورتی که پرسش مطرح شده مبهم باشد سیستم پرسش و پاسخ، تعاملی را با کاربر انجام می‌دهد و پس از رفع ابهام پاسخ آن را تعیین می‌کند؛
- اگر پرسش مطرح شده دارای بخش حذف شده‌ای باشد، (مانند پرسش چرا؟) سیستم با استفاده از کلمات کلیدی سوال قبل، آن‌را تکمیل می‌نماید و پس از بازسازی در صورت تایید کاربر پاسخ آن را تعیین می‌نماید؛
- اگر پرسش مطرح شده دارای ضمیر باشد، سیستم سعی می‌کند ضمایر را با مراجع مناسب جایگزین نموده و سپس سوال اصلاح شده را به کاربر بازگردانده و از او می‌خواهد تایید کند و در صورتی که کاربر اصلاحات را تایید نماید، سیستم پاسخ پرسش بازسازی شده را تعیین می‌نماید؛
- اگر پاسخگویی به پرسش نیاز به تعامل نداشته باشد، سیستم پاسخ آن را تعیین می‌کند؛
- چرخه فوق تا اتمام مکالمه از سوی کاربر ادامه می‌یابد [۳۹،۳۸].

۴-۶ پژوهش‌های مرتبط در حوزه پرسش و پاسخ تعاملی

هدف از این تحقیق، طراحی و پیاده‌سازی یک سیستم پرسش و پاسخ تعاملی مستقل از زبان است که بتواند با بکارگیری تکنیک‌های آماری، ابهام در پرسش را برطرف نموده و امکان طرح سوالات مرتبط را به کاربر بدهد. سیستم‌های پرسش و پاسخ تعاملی موجود عمدتا یا از تجزیه و تحلیل‌های معنایی و گرامری استفاده می‌کنند که آن‌ها را محدود به زبان خاص می‌نماید و یا نیازمند آموزش با پیکره بزرگی از دادگان از طریق بکارگیری تکنیک‌های یادگیری ماشین می‌باشند. در این بخش الگوریتم‌های استفاده شده در این سیستم‌ها تشریح می‌شود.

۴-۶-۱ روش‌های پاسخگویی به سوالات دارای ضمیر

۱) روش جستجوی پاسخ بدون در نظر گرفتن ضمیر: در این روش که ساده‌ترین روش است، سیستم ضمیر را از پرسش کاربر حذف نموده و سپس با فرض اینکه وجود ضمیر در یک پرسش نشانه ارتباط آن پرسش با پرسش قبلی است، پاسخ پرسش دارای ضمیر را در میان اسنادی که جهت پاسخگویی به سوال قبلی بازگردانده شده‌اند جستجو می‌کند. این روش زمانی که دو پرسش مطرح شده موضوع یکسانی داشته باشند، از کارایی مناسبی برخوردار است. در حالتی که موضوع دو پرسش متفاوت باشد، بکارگیری این روش می‌تواند منجر به بازگرداندن پاسخ نادرست شود [۴۰].

۲) روش جایگزینی ضمیر با موضوع مکالمه: در این روش با استفاده از الگوریتم‌هایی مانند تئوری مرکزیت[1]، موضوع اصلی مکالمه تعیین و جایگزین ضمیر می‌شود. گرچه در بسیاری از موارد مرجع ضمیر موضوع اصلی مکالمه است، بعضی از ضمایر به موردی غیر از موضوع مکالمه اشاره می‌کنند و همین امر می‌تواند به بازگرداندن پاسخ نادرست منجر شود [۴۱].

[1] Centering Teory

۳) روش تعیین مرجع ضمیر با الگوریتم‌های مبتنی بر گرامر: در این روش که با فرض وجود یک درخت پارس کامل و صحیح صورت می‌گیرد، درخت برای یافتن مرجع پیمایش می‌شود و محدودیت‌ها روی مراجع احتمالی یافت شده اعمال می‌گردند. بر اساس محدودیت‌ها، تعدادی از کاندیداها حذف و سپس بر اساس بعضی عوامل تعیین کننده، بهترین عبارت اسمی به عنوان مرجع از بین باقیمانده کاندیداها انتخاب و جایگزین می‌شود [۴۲]. از جمله محدودیت‌هایی که در این روش در نظر گرفته می‌شود، می‌توان به موارد زیر اشاره کرد:

- سازگاری مرجع با ضمیر از نظر تعداد
- سازگاری مرجع با ضمیر از نظر شخص و وضعیت
- سازگاری مرجع با ضمیر از نظر جنسیت
- سازگاری مرجع با ضمیر از نظر نحوی

هم‌چنین در نظر گرفتن عوامل زیر می‌تواند به انتخاب مرجع مناسب کمک کند:

- تازگی: موجودیت‌های جدید مرجع‌های محتمل‌تری هستند؛
- نقش گرامری در جمله: موجودیت‌هایی که در نقش فاعلند مرجع‌های محتمل‌تری هستند؛
- تعداد دفعات تکرار: موجودیت‌هایی که بیشتر تکرار شده‌اند مرجع‌های محتمل‌تری هستند؛
- فاصله مرجع تا ضمیر: موجودیت‌هایی که با ضمیر فاصله کمتری دارند، مرجع‌های محتمل‌تری هستند؛
- موازی‌کاری[11]: ساختار یکسان در دو جمله به انتخاب مرجع مناسب کمک می‌نماید.

مثالی از رویکرد مبتنی بر گرامر، الگوریتم Hobbs است که یک جستجوی اول سطح چپ به راست انجام می‌دهد و به مراجع احتمالی که نزدیک‌تر هستند برتری می‌دهد. برخی از معایب روش‌های مبتنی بر گرامر به شرح زیر است:

- در رویکرد مبتنی بر نحو از روابط معنایی استفاده نمی‌شود و همین امر می‌تواند یافتن مرجع

[11] Parallel

مناسب برای کلمات مخفف و نیز کلمات و عبارات معادل را به شدت محدود نماید.

- هنگامی که چند مرجع احتمالی با شرایط مناسب وجود داشته باشد، سیستم ممکن است نتواند مرجع صحیح را تشخیص دهد.
- اگر درخت پارس نادرست باشد، مرجع به درستی تشخیص داده نمی‌شود.

۴) روش تعیین مرجع با الگوریتم‌های مبتنی بر یادگیری ماشین: در روش تعیین مرجع با الگوریتم‌های مبتنی بر یادگیری ماشین، دانش مورد نیاز جهت تعیین مرجع مناسب با استفاده از الگوریتم‌های یادگیری و داده‌های آموزشی به دست می‌آید. در این روش برای آموزش سیستم از یک پایگاه دادگان که در آن زوج‌های "عبارت اسمی مرجع و ضمیر" نشانه‌گذاری شده‌اند استفاده می‌شود. متن آموزشی باید قبلا توسط چندین پیمانه پردازش زبان طبیعی پیش‌پردازش شده باشد به طوریکه تمامی ضمایر و مراجع آن‌ها مشخص باشد. سپس برای هر زوج مرجع-ضمیر، یک بردار ویژگی تشکیل شده و در قالب نمونه آموزشی به الگوریتم یادگیری داده می‌شود. بر اساس این بردار ویژگی، الگوریتم‌های یادگیری درباره تشخیص مرجع بودن یا نبودن عبارت اسمی مورد بررسی و یا احتمال مرجع بودن آن آموزش می‌بینند. شیوه‌های مختلف ارائه شده جهت آموزش سیستم برای تعیین مراجع از انواع متفاوتی از ویژگی‌ها (لغوی، نحوی و معنایی) و اندازه‌های متفاوتی از بردارهای ویژگی استفاده کرده‌اند. الگوریتم بیزین و درخت تصمیم‌گیری اولین تکنیک‌های بکار رفته یادگیری ماشین در تعیین مرجع می‌باشند. در الگوریتم بیزین پس از آموزش عامل یادگیر، تمام مراجع احتمالی برای هر یک از ضمایر مشخص شده و احتمال مرجع بودن آن‌ها برای ضمیر مورد بررسی تعیین می‌شود. در نهایت محتمل‌ترین گزینه به عنوان مرجع ضمیر انتخاب و جایگزین می‌شود [۴۳،۴۲].

۴-۶-۱-۱ کارهای انجام شده جهت جایگزینی ضمایر با مراجع در متون فارسی

محققان موسوی و قاسم ثانی از یک روش یادگیری ماشین مبتنی بر رتبه‌بندی برای تعیین مراجع ضمایر فارسی استفاده کرده‌اند [۴۴]. آن‌ها از پایگاه دادگان PCAC-2008 به عنوان مجموعه آموزش استفاده کردند و با بکارگیری روش رتبه‌بندی بیشینه بی‌نظمی[1]، مراجع احتمالی یک ضمیر را رتبه‌بندی کرده تا مرجع با بهترین رتبه به عنوان مرجع نهایی انتخاب شود. با وجود راحت بودن پیاده‌سازی، این روش معایبی نیز دارد که از آن جمله می‌توان به کارایی پایین آن به دلیل عدم وجود دستور زبان واحد در فارسی و نیز عدم وجود جنسیت در ضمایر فارسی که نتیجه آن از دست دادن یکی از عوامل اصلی شفاف‌سازی ضمایر می‌باشد اشاره نمود.

در تحقیقی که توسط فلاحی و شمس فرد انجام شده است، از یک مجموعه قوانین برای پیدا نمودن مراجع ضمایر استفاده شده است [۴۵]. در سیستم طراحی شده آن‌ها دو مرحله اصلی زیر وجود دارد:

- مرحله پردازش: در این مرحله پس از تشخیص بازه کلمات، از شناسایی و برچسب زدن اجزای واژگانی کلام استفاده می‌شود.
- مرحله شفاف‌سازی[2]: در این مرحله از خروجی مرحله قبلی استفاده شده و مرجع هر ضمیر با بکارگیری قوانین زیر مشخص می‌شود:
  - اگر ضمیر در بخش فاعلی جمله دوم باشد، مرجع به احتمال زیاد در بخش فاعلی جمله اول است.
  - اگر ضمیر در بخش مفعولی جمله دوم باشد، مرجع به احتمال زیاد در بخش مفعولی جمله اول است.
  - اگر ضمیر "او" در جمله وجود داشته باشد، مرجع آن یک اسم عاقل در جمله قبل است.

[1] Maximum entropy eanker

[2] Resolution

- اگر ضمیر "آن" در جمله وجود داشته باشد، مرجع این ضمیر یک اسم غیر عاقل در جمله قبل است.
- اگر ضمیر "آن‌ها" در جمله وجود داشته باشد، مرجع این ضمیر یک اسم جمع یا مجموعه‌ای از اسامی که به وسیله "و" به هم متصل شده‌اند می‌باشد.
- اگر ضمیر "ما" در جمله وجود داشته باشد، مرجع "من و ..." است.
- اگر ضمیر "اینجا" یا "آنجا" در جمله وجود داشته باشد، مرجع آن یک مکان در جملات قبلی است.

۴-۶-۱-۲ معایب الگوریتم‌های موجود برای تعیین مرجع ضمایر

از معایب الگوریتم‌های تعیین مرجع ضمایر می‌توان به موارد زیر اشاره کرد:

۱) بسیاری از ویژگی‌هایی که در روش‌های زبان‌شناختی و روش‌های یادگیری ماشین جهت تعیین مرجع مناسب استفاده می‌شوند، منحصر به زبان خاصی می‌باشند و عدم امکان استفاده از این ویژگی‌ها در دیگر زبان‌ها، کارایی روش‌های موجود را در تعیین مراجع کاهش می‌دهد. اکنون تفاوت‌های موجود بین ویژگی‌های ضمایر زبان فارسی و زبان انگلیسی که می‌تواند باعث ناکارامدی استفاده از روش‌های تشخیص مراجع در زبان انگلیسی برای زبان فارسی شود تشریح می‌گردد:

- یکی از ویژگی‌های مهمی که در تعیین مراجع ضمایر انگلیسی استفاده می‌شود، تطبیق جنس ضمیر و مرجع آن است. به علت اینکه ضمایر در زبان فارسی دارای ویژگی جنس نیستند، این ویژگی در فرایند تعیین مرجع ضمایر فارسی کاربردی نخواهد داشت.
- ویژگی مهم دیگری که در تعیین مرجع یک ضمیر در زبان انگلیسی بکار می‌رود ویژگی مطابقت عددی است (ضمایر مفرد به یک عبارت اسمی مفرد و ضمایر جمع به یک عبارت اسمی جمع اشاره دارند). در مقابل در زبان فارسی ضمایر مفرد می‌توانند به یک عبارت اسمی جمع اشاره داشته باشند و گاهی نیز جهت احترام به اشخاص به جای ضمیر مفرد از ضمیر جمع

استفاده می‌شود. به همین دلیل ویژگی مطابقت عددی نیز در فرایند تعیین مرجع ضمایر فارسی کم‌اثر می‌شود.

- نقش ضمیر در جمله فارسی تاثیری بر شکل آن ندارد (ضمیر فارسی در نقش‌های فاعلی، مفعولی و ملکی به یک شکل استفاده می‌شود). به همین دلیل نمی‌توان از نقش ضمیر در جمله فارسی به عنوان یک ویژگی برای تعیین مرجع استفاده نمود.

۲) تعیین عبارات اسمی و استخراج ویژگی‌های نحوی آنها که بخش مهمی از ویژگی‌های عبارات را تشکیل می‌دهند، نیازمند استفاده از تجزیه‌گر و پارسر خاص آن زبان است و این امر روش‌های بکار رفته را وابسته به زبان می‌نماید. بنابراین استفاده از ویژگی‌های نحوی برای تعیین مرجع ضمایر یک زبان نیازمند توسعه پارسر مناسب که مختص آن زبان است می‌باشد.

۳) با توجه به پیچیدگی‌های زبان، موارد زیادی وجود دارد که تعیین مرجع برای ضمایر بکار رفته در جمله به آسانی امکان‌پذیر نیست و بهترین روش برای تشخیص مرجع ضمیر توسط سیستم، برقراری تعامل با کاربر و درخواست تعیین مرجع ضمیر توسط خود کاربر می‌باشد. در این‌گونه موارد الگوریتم‌های مبتنی بر پردازش نحوی زبان طبیعی و نیز الگوریتم‌های مبتنی بر یادگیری ماشین نمی‌توانند به صورت منطقی و مستدل مرجع ضمیر را تشخیص دهند و هرگونه انتخاب مرجع برای ضمیر توسط آن‌ها صرفا به صورت اتفاقی خواهد بود. مثال‌های زیر نمونه‌هایی از جملاتی هستند که تشخیص مرجع ضمیر توسط الگوریتم‌های زبان‌شناختی یا یادگیری ماشین امکان‌پذیر نیست و تعیین مرجع باید به روش مستقل از زبان و با تایید خود کاربر صورت پذیرد:

- مثال اول (ابهام): "برگ از درخت افتاد، زیرا آن خشک شده بود."
- مثال دوم (ابهام): "پسر به پیرمرد کمک کرد، زیرا باعث خوشحالی او می‌شد."
- مثال سوم (نیاز به فهم معنایی): "باران پشت‌بام خانه را تخریب کرد، زیرا آن فرسوده بود."

"باران پشت‌بام خانه را تخریب کرد، زیرا آن شدید بود."

- مثال چهارم (حذف مرجع): "من به بیمارستان رفتم و آنها به من گفتند که من باید به خانه

بروم و استراحت کنم."

در مثال اول و دوم در تعیین مرجع ابهام وجود دارد، زیرا دو عبارت اسمی به عنوان مرجع بالقوه ضمیر وجود دارند. در مثال سوم، تشخیص مرجع ضمیر فقط با توجه به معنای جمله امکان‌پذیر است و دستور زبان در این مورد کمکی نمی‌کند. در مثال چهارم، مرجع مشخصی برای ضمیر در جمله وجود ندارد و باید از خود کاربر برای تعیین مرجع کمک گرفت.

۴-۶-۲ الگوریتم تشخیص پرسش‌های حذفی و پاسخگویی به آن‌ها

در پرسش‌های حذفی بخشی از جمله مانند فعل با توجه به جمله قبلی حذف می‌شود. روش‌های مبتنی بر گرامر با استفاده از پارسرهای موجود برای یک زبان، فعل جمله را جستجو می‌کند و در صورتی‌که فعل وجود نداشته باشد، پرسش از نوع حذفی خواهد بود. جهت بازسازی پرسش حذفی، کلمات کلیدی پرسش قبل به آن افزوده شده و سپس جستجوی پاسخ بر اساس مجموعه کلمات کلیدی به‌روز شده انجام می‌شود. اشکال این روش وابستگی آن به زبان خاص می‌باشد [۴۰،۳۸].

۴-۶-۳ الگوریتم خطایابی املایی

با گسترش استفاده از اینترنت، متون الکترونیکی و سیستم‌های پرسش و پاسخ، وجود الگوریتم‌های تشخیص خطای املایی و ارائه کلمات پیشنهادی مناسب دارای اهمیت روزافزونی شده‌اند. وابستگی الگوهای خطا به یک زبان شناسایی آن را دشوار و زمان‌بر می‌کند. خطاهای املایی انجام شده توسط کاربران انواع مختلفی دارد که عبارتند از [۴۶]:

- خطای عدم آگاهی کاربر: کاربر تصور می‌کند که واژه را به طور صحیح نوشته است، اما واژه از نظر املایی نادرست می‌باشد.
- خطای حذف: در این نوع خطا یک یا چند حرف واژه به صورت ناخواسته حذف می‌گردد.

- خطای درج: در این نوع خطا یک یا چند حرف به صورت ناخواسته به کلمه اضافه می‌شود.
- خطای جابجایی: در این نوع خطا دو حرف کنار هم به صورت ناخواسته جابجا وارد می‌شوند.
- خطای چسبیدن: در این نوع خطا کاربر ناخواسته بین دو کلمه فاصله نمی‌گذارد و دو کلمه به هم متصل می‌شود.
- خطای جانشینی: در این نوع خطا کاربر به صورت اشتباه کلید دیگری را به جای کلید مورد نظر لمس می‌کند و بنابراین حرف دیگری به جای حرف مورد نظر وارد می‌شود. این خطا معمولا در مورد حروف کنار هم در صفحه کلید اتفاق می‌افتد.

برای شناسایی و تصحیح رشته‌هایی که دارای اشتباه املایی هستند و به آن شکل در لغت‌نامه وجود ندارند، روش‌های مختلفی وجود دارد که تعدادی از آنها شرح داده می‌شود.

۴-۶-۳-۱ روش‌های تشخیص خطا

- تحلیل N-گرام: N-گرام‌ها زیردنباله‌هایی از رشته‌ها و کلمات می‌باشند که چند حرفی بوده و N معمولا یک، دو یا سه می‌باشد. در روش‌هایی که از N-گرام جهت شناسایی خطای املایی استفاده می‌شود، هر یک از N-گرام‌های موجود در متن ورودی انتخاب و در یک جدول از قبل آماده شده مورد جستجو قرار می‌گیرد. در این روش رشته‌هایی که دارای N-گرام‌های بدون تکرار یا کم تکرار هستند (برای مثال آشز یا شخس) به عنوان رشته‌هایی که احتمالا دارای خطای املایی هستند در نظر گرفته می‌شوند. در این روش برای تولید جدول N-گرام‌ها از لغت‌نامه یا مجموعه بزرگی از متون استفاده می‌شود.
- روش مبتنی بر لغت‌نامه: در این روش کلمات وارد شده توسط کاربر در لغت‌نامه جستجو می‌شود و چناچه کلمه در لغت‌نامه وجود نداشته باشد، به عنوان خطا علامت‌گذاری می‌شود. حجم زیاد لغت‌نامه‌ها می‌تواند سرعت پاسخگویی را کاهش دهد و به همین دلیل معمولا از جدول‌های درهم با هدف افزایش سرعت جستجو استفاده می‌شود [۴۶].

۴-۶-۳-۲ روش‌های تصحیح خطا

- روش کمترین فاصله تصحیح: در این روش تغییراتی در کلمه دارای خطای املایی داده می‌شود و کلمات جدید ایجاد شده در فهرست واژگان صحیح آن زبان که می‌تواند یک لغت‌نامه باشد جستجو می‌گردد تا واژگان صحیح احتمالی پیدا شوند. سپس از بین کلمات استخراج شده، محتمل‌ترین گزینه یا گزینه‌ها برای جایگزینی با کلمه دارای خطای املایی مشخص و پیشنهاد می‌شود.
- روش کلید مشابه: در این روش با توجه به ترتیب حروف روی صفحه کلید و مشابهت‌های بین آن‌ها، کلمات در گروه‌های مشابهی قرار داده می‌شوند تا بتوان از آنها در انتخاب واژه جایگزین مناسب برای کلمه دارای خطای املایی استفاده نمود.
- روش آماری: در روش آماری که از کارآمدترین روش‌ها می‌باشد، فهرستی از گزینه‌های احتمالی تولید و رتبه‌بندی می‌شود. بر طبق قانون بیز احتمال مناسب بودن واژه جایگزین X از فرمول زیر محاسبه می‌شود:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \qquad (۱-۴)$$

در این فرمول X واژه جایگزین احتمالی، Y واژه نادرست بکاررفته توسط کاربر و $P(X|Y)$ احتمال مناسب بودن گزینه جایگزین X را نشان می‌دهد. $P(Y|X)$ احتمال وقوع خطای املایی که منجر به استفاده از Y به جای X شده است را نشان می‌دهد. $P(X)$ نیز احتمال استفاده از لغت X است که برای محاسبه آن، تعداد نرمال شده آن در پیکره در نظر گرفته می‌شود [۴۷،۴۶].

۴-۶-۳-۳ معایب بکارگیری الگوریتم‌های خطایابی در پرسش و پاسخ

۱) شناسایی و تصحیح خطاهای املایی بر اساس لغت‌نامه و یا پیکره بزرگ زبان زمان‌بر بوده و همچنین استفاده از سیستم را به یک زبان خاص محدود می‌کند.

۲) الگوریتم‌های خطایابی بافت جمله را مورد توجه قرار نمی‌دهند و به همین دلیل کلمه‌ای که با

توجه به جمله دارای خطای املایی است ولی ظاهری درست دارد، شناسایی نخواهد شد. مثلا واژه "اسب" در جمله زیر درست به شمار می‌آید: "هوا گرم اسب."

۳) هنگام استفاده از سیستم پرسش و پاسخ با رویکرد آماری، بکار رفتن کلماتی در پرسش که با وجود صحیح بودن از نظر املا در منابع داده ذکر نشده‌اند، کارایی سیستم را پایین می‌آورد. در مواردی ممکن است به جای کلمه اصلی به کار رفته در پرسش کاربر، مترادف آن در منابع داده وجود داشته باشد اما توسط سیستم تشخیص داده نشود.

۴-۷ ارزیابی سیستم‌های پرسش و پاسخ تعاملی

ارزیابی سیستم‌های پرسش و پاسخ تعاملی به منظور تعیین و ارتقای کارایی آن‌ها از اهمیت زیادی برخوردار است. با این وجود هنوز روش استاندارد و مخصوصی برای ارزیابی این سیستم‌ها ارائه نشده است و ارزیابی معمولا با استفاده از روش‌های ارزیابی سیستم‌های پرسش و پاسخ اتوماتیک و نیز سیستم‌های مکالمه محور صورت می‌گیرد. در ارزیابی سیستم‌های پرسش و پاسخ تعاملی علاوه بر ارزیابی کمی از ارزیابی کیفی نیز استفاده می‌شود که نیازمند مشارکت کاربران در فرایند ارزیابی برای تعیین میزان موفقیت تعامل بین سیستم و کاربر می‌باشد [۱۸]. در ادامه روش‌های ارزیابی کمی[1] و کیفی[2] که در سیستم‌های پرسش و پاسخ تعاملی مورد استفاده قرار می‌گیرند، تشریح می‌شوند.

۴-۷-۱ ارزیابی کمی

ارزیابی کارایی سیستم‌های پرسش و پاسخ از نظر صحت و دقت پاسخگویی آن‌ها با توجه به نوع سیستم اندکی متفاوت است. برای ارزیابی سیستم‌هایی که به منظور پاسخگویی کوتاه به سوالات مبتنی بر حقایق طراحی شده‌اند، یک مجموعه استاندارد پرسش و پاسخ تهیه شده و ارزیابی سیستم

[1] Quantitative criteria

[2] Qualitative criteria

بر اساس میزان مطابقت پاسخ‌های سیستم با این مجموعه صورت می‌گیرد. این روش ارزیابی برای سیستم‌هایی که به منظور پاسخگویی به سوالات پیچیده طراحی شده‌اند چندان مناسب نیست، زیرا پاسخ صحیح یک سوال می‌تواند به شکل‌های مختلفی بیان شود [۳۹،۳۸]. ضمن اینکه بعضی از این سوالات ممکن است بیش از یک پاسخ صحیح داشته باشند. ارزیابی این‌گونه سیستم‌ها معمولا به صورت دستی و توسط شخص صورت می‌گیرد. برای نمونه می‌توان به ارزیابی سیستم‌های پرسش و پاسخ شرکت کننده در کنفرانس سالانه بازیابی متون (TREC) اشاره نمود. با توجه به ماهیت سیستم و نحوه بازگرداندن پاسخ، ارزیابی کمی بر اساس یک یا چند معیار ارزیابی که در بخش ۳-۳-۴ شرح داده شده است انجام می‌شود.

۴-۷-۲ ارزیابی کیفی

هدف از ارزیابی کمی سیستم‌های پرسش و پاسخ تعاملی، تعیین میزان صحت پاسخ بازگردانده شده توسط این سیستم‌ها می‌باشد. این ارزیابی، اطلاعات کافی درباره کیفیت تعامل سیستم با کاربر و اینکه آیا تعامل موفقیت‌آمیز خاتمه یافته است یا خیر، ارائه نمی‌کند. به همین دلیل علاوه بر ارزیابی کمی، این سیستم‌ها از نظر کیفی نیز مورد سنجش و ارزیابی قرار می‌گیرند تا کیفیت تعامل سیستم با کاربر و میزان رضایتمندی کاربر تعیین شود. جهت ارزیابی کیفیت تعامل، معمولا پرسشنامه‌ای تهیه شده و در اختیار کاربران قرار می‌گیرد تا با تکمیل آن میزان رضایت خود را از کیفیت تعامل اعلام کنند [۱۸]. هاراباگیو و همکارانش [۴۸] بر اندازه‌گیری سه ویژگی از سیستم‌های پرسش و پاسخ تعاملی به وسیله پرسشنامه تاکید می‌کنند:

۱) کارآمدی[1]: تعداد پرسش‌هایی که کاربر باید برای دستیابی به اطلاعات مورد نظر مطرح کند

۲) تاثیرگذاری[2]: صحت پاسخ‌های ارائه شده توسط سیستم

[1] Efficiency

[2] Effectiveness

۳) رضایت عمومی کاربر از عملکرد سیستم

کوارترونی و ماناندهار [۳۷] پرسش‌های زیر را جهت ارزیابی سیستم پرسش و پاسخ تعاملی خود، مطرح کردند و از کاربران خواستند با دادن امتیازی بین یک (حداقل امتیاز) تا پنج (حداکثر امتیاز) کیفیت تعامل را اندازه‌گیری نمایند:

۱) به چه میزان توانستید اطلاعات مورد نظر خود را از سیستم به دست آورید؟

۲) آیا فکر می‌کنید سیستم در درک منظور شما موفق بوده است؟

۳) آیا دسترسی به اطلاعات موردنظرتان برای شما ساده بوده است؟

۴) آیا درخواست سیستم در بیان مجدد سوال با کلمات یا ساختار جدید به نظرتان منطقی بوده است؟

۵) آیا کار با این سیستم بگونه‌ای بوده است که در آینده مجددا از این سیستم استفاده نمایید؟

۶) آیا سرعت سیستم در پاسخگویی به سوالات شما مناسب بوده است؟

۷) در کل، میزان رضایت خود از سیستم را اعلام نمایید؟

پرسش‌های (۱) و (۲) عملکرد سیستم، پرسش‌های (۳) و (۴) مشکلات تعامل، پرسش (۶) سرعت پاسخگویی و پرسش‌های (۵) و (۷) رضایت کلی کاربر از سیستم را ارزیابی می‌کنند.

از آنچه گفته شد می‌توان نتیجه گرفت ارزیابی سیستم‌های پرسش و پاسخ تعاملی نیازمند هر دو معیار کمی و کیفی است تا هم کارایی سیستم در ارائه پاسخ‌های صحیح و دقیق مورد ارزیابی قرار گیرد و هم میزان رضایتمندی کاربر و به عبارتی کارایی کیفی سیستم مشخص گردد.

فصل پنجم

روش پیشنهادی و بررسی نتایج و ارزیابی

۵-۱ روش پیشنهادی

هدف از این تحقیق، طراحی یک سیستم پرسش و پاسخ است که بتواند مستقل از زبان و حوزه دانش، نیاز اطلاعاتی کاربران را برطرف نماید. این سیستم باید بتواند با در اختیار داشتن پایگاه دادگان مناسب هر زبان به سوالات مطرح شده به آن زبان پاسخ دهد و با افزودن سطح تعامل، امکان پاسخگویی به دنباله‌ای از سوالات مرتبط و مبهم را فراهم نماید. باید توجه داشت که سیستم‌های پرسش و پاسخ تعاملی موجود عمدتا یا از تجزیه و تحلیل‌های معنایی و گرامری استفاده می‌کنند که آن‌ها را محدود به زبان خاص می‌نماید و یا نیازمند آموزش با پیکره بزرگی از دادگان از طریق بکارگیری تکنیک‌های یادگیری ماشین می‌باشند. ویژگی که سیستم طراحی شده را از سیستم‌های موجود متمایز می‌کند، استخراج دانش نهفته در متون با استفاده از رویکرد آماری جهت پاسخگویی و برقراری تعامل است.

تکنیک‌های آماری استفاده شده، علاوه بر ایجاد یک سیستم مستقل از زبان سرعت جستجو را افزایش می‌دهد. با این وجود کاهش دقت پاسخگویی با توجه به عدم استفاده از دانش زبان‌شناختی امری محتمل است. از این‌رو سیستم طراحی شده به جای پاسخ دقیق، یک قطعه از متون را بازمی‌گرداند تا احتمال ارائه پاسخ مناسب افزایش یابد. توانایی سیستم در بازگرداندن قطعه‌ای از متون، آن را قادر می‌سازد که هم به سوالاتی درباره حقایق کوتاه مانند "کجا"، "کی" و "چه کسی" و هم به سوالات پیچیده‌تری مانند "چرا" و "چگونه" که به پاسخ‌های طولانی‌تری نیاز دارند پاسخ دهد. سیستم طراحی شده همچنین می‌تواند با ارائه اطلاعات بیشتر کاربر را از طرح برخی از سوالات بی‌نیاز نماید [۳۸].

روش‌های پیشنهاد شده در این تحقیق به دو پرسش کلی پاسخ می‌دهند: ۱) چگونه سیستم به کاربر کمک کند تا پرسش خود را بهتر مطرح کند. ۲) چگونه پس از طرح پرسش، سیستم با رفع ابهامات و خطاهای احتمالی پاسخ مناسب‌تری به کاربر بازگرداند.

برای برآوردن اهداف فوق، سیستم طراحی شده دارای ۶ ماژول زیر است:

۱) طراحی ماژول پیشنهاد دهنده عبارات پرتکرار در زمان طرح پرسش

۲) طراحی ماژول تصحیح خطاهای املایی

۳) طراحی ماژول تعیین مراجع مرتبط با ضمایر

۴) طراحی ماژول بازسازی سوالات حذفی

۵) ماژول بازیابی قطعه متون با تکنیک‌های آماری – JIRS

۶) طراحی ماژول پیشنهاد دهنده پرسش‌های پرتکرار

ماژول‌های ۱ و ۶ جهت طرح بهتر پرسش و بقیه جهت پاسخگویی مناسب طراحی شده‌اند.

۵-۲ معماری سیستم پیشنهادی

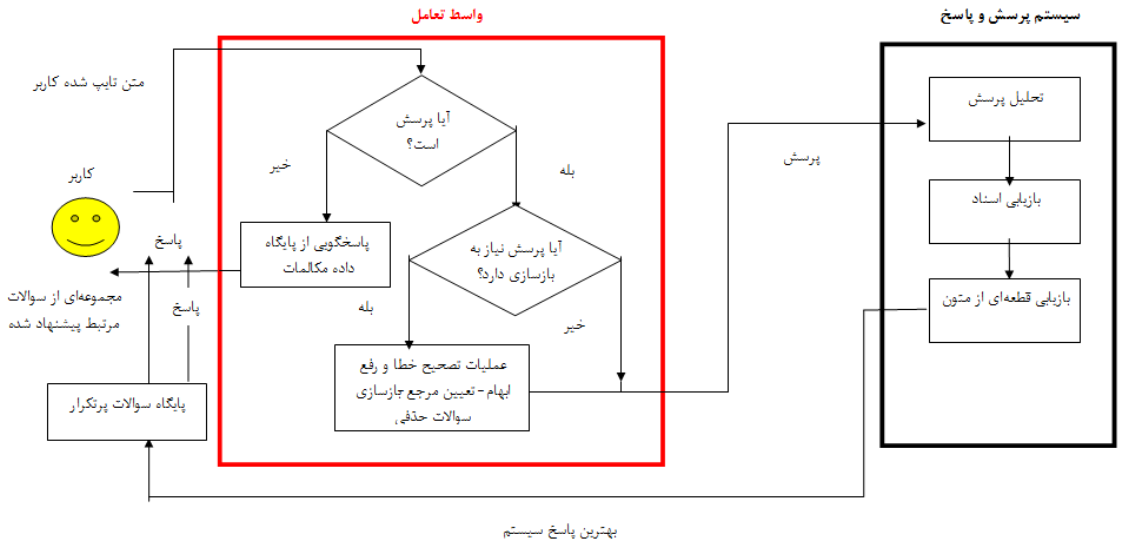
سیستم پرسش و پاسخ طراحی شده به شکل زیر عمل می‌کند:

- کاربر پرسش خود را مطرح می‌کند؛ در این فرایند ماژول پیشنهاد دهنده عبارات در پرسش، به کاربر کمک می‌نماید تا پرسش خود را به شکل مناسب‌تری طرح نماید.
- پرسش ورودی بررسی می‌شود تا مشخص شود آیا پاسخگویی به آن نیاز به تعامل با کاربر دارد یا خیر.
- اگر پرسش مطرح شده دارای خطای املایی و یا کلمات ناشناخته باشد، ماژول تصحیح خطای املایی با رویکرد تعاملی خود، ابهام را برطرف می‌نماید.
- اگر پرسش مطرح شده دارای بخش حذف شده باشد، ماژول بازسازی سوالات حذفی پرسش کاربر را تکمیل می‌نماید.
- اگر پرسش مطرح شده دارای ضمیر باشد، ماژول تعیین مراجع با رویکرد تعاملی مرجع ضمیر را جایگزین می‌نماید.
- در مرحله بعد، ابتدا با استفاده از TF*IDF سند مناسب از میان مجموعه اسناد انتخاب می‌شود

و سپس سیستم با استفاده از الگوریتم JIRS ، قطعه‌ای از متون را که به احتمال زیاد شامل پاسخ صحیح است تعیین و به کاربر نمایش می‌دهد.

- ماژول پیشنهاد دهنده پرسش‌های پرتکرار، فهرستی از پرسش‌های مرتبط با پرسش فعلی را به کاربر نمایش می‌دهد تا در صورت تمایل آن را جهت ادامه گفتگو انتخاب نماید.
- چرخه فوق تا اتمام مکالمه از سوی کاربر ادامه می‌یابد.

شکل ۵-۱ معماری سیستم پیشنهاد شده را نمایش می‌دهد.



شکل ۵-۱ معماری سیستم پرسش و پاسخ طراحی شده

۵-۲-۱ پیشنهاد عبارات پرتکرار در پرسش کاربر

سیستم پرسش و پاسخ طراحی شده، مستقل از گرامر و معنا بوده و فقط از تکنیک‌های مبتنی بر آمار جهت پاسخگویی به پرسش کاربر استفاده می‌کند و بنابراین استفاده از عبارات اسمی صحیح و طرح سوال به شکل مناسب توسط کاربر از اهمیت بالایی برخوردار است. در سیستم پرسش و پاسخ طراحی شده، ماژولی افزوده شده است که هوشمندانه به کاربر کمک می‌کند پرسش خود را با استفاده از عبارات اسمی مطرح کند که پرتکرار بوده و ارزش بیشتری دارند. این ماژول قادر است با در نظر

گرفتن پرسش‌هایی که تاکنون کاربران از این سامانه مطرح کرده‌اند و نیز مجموعه اسنادی که در فرایند پاسخگویی مورد استفاده قرار می‌گیرند، N-گرام‌های با ارزش را جهت طرح پرسش مناسب به کاربر پیشنهاد نماید. در روش پیشنهاد شده پس از تایپ دومین کاراکتر هر کلمه از پرسش، سیستم عبارات سه‌کلمه‌ای موجود در مجموعه متون و پرسش‌های مطرح شده قبل را که با این دو کاراکتر شروع می‌شوند استخراج نموده و با توجه به تعداد تکرار هر یک از عبارات در مجموعه متون و پرسش‌های قبل، وزنی به آن‌ها اختصاص می‌دهد. سپس سیستم ۵ عبارت دارای بیشترین ارزش را به ترتیب اولویت به کاربر پیشنهاد می‌نماید تا در صورت تمایل یکی از آن‌ها را انتخاب نماید. چنانچه تعداد عبارات سه‌کلمه‌ای که با دو کاراکتر تایپ شده آغاز می‌شوند کمتر از ۵ باشد، بقیه پیشنهادات از میان عبارات دوکلمه‌ای و سپس عبارات یک‌کلمه‌ای انتخاب می‌شوند. به عنوان مثال فرض کنید کاربر قصد مطرح کردن پرسش زیر را دارد: "آیا دانشجو پس از مشروط شدن در دو ترم متوالی اخراج می‌شود؟"

در این سوال اولین کلمه‌ای که کاربر تایپ می‌کند "آیا" است. با تایپ "آی" سیستم ۳ عبارت سه‌کلمه‌ای "آیا امکان فارغ التحصیل"، "آیا حداقل معدل" و "آیا انتقال دانشجو" را در مجموعه متون و پرسش‌های قبلی کاربران پیدا می‌کند که همگی با "آی" شروع شده‌اند. در میان آن‌ها عبارت "آیا امکان فارغ‌التحصیل" ۲ بار تکرار شده و ارزش بیشتری دارد. در نتیجه این عبارت اولین پیشنهاد سیستم به کاربر خواهد بود و دو عبارت "آیا حداقل معدل" و "آیا انتقال دانشجو" هر کدام با یک تکرار، پیشنهادات بعدی می‌باشند. اما امکان ارائه دو پیشنهاد دیگر باقیست. از آنجا که عبارت سه‌کلمه‌ای دیگری که با "آی" شروع شود در مجموعه متون و پرسش‌ها وجود ندارد، برای ارائه دو پیشنهاد دیگر عبارات دوکلمه‌ای که با "آی" شروع می‌شوند استخراج می‌شوند که در این مورد عبارات "آینده است"، "آینده دارد" و "آینده می‌باشد" هستند. از آنجا که عبارات "آینده دارد" و "آینده می‌باشد" دو بار تکرار شده‌اند، ارزش بیشتری نسبت به عبارت "آینده است" داشته و به کاربر پیشنهاد می‌شوند. لازم به ذکر است اگر عبارت دوکلمه‌ای که با "آی" شروع شود وجود نداشته باشد،

در مرحله بعد عبارات یک‌کلمه‌ای که با "آی" آغاز شده، استخراج می‌شوند و از بین آن‌ها عبارات یک‌کلمه‌ای با بیشترین ارزش پیشنهاد می‌شوند.

در ادامه با تایپ دو کاراکتر "دان" مربوط به کلمه "دانشجو"، سیستم عبارات سه‌کلمه‌ای را که با "دا" شروع می‌شود استخراج نموده و ۵ عبارت دارای بیشترین ارزش را به کاربر پیشنهاد می‌نماید. در صورتی که تعداد عبارات سه‌کلمه‌ای کمتر از ۵ باشد، عملیات فوق برای عبارات دوکلمه‌ای و سپس یک‌کلمه‌ای تکرار می‌شود. شکل ۵-۲ نمونه‌ای از پیشنهادات ارائه شده در زمان طرح پرسش را نشان می‌دهد.

شکل ۵-۲ نمونه‌ای از پیشنهادات ارائه شده در زمان طرح پرسش

۵-۲-۲ تصحیح خطاهای املایی و رفع ابهام

خطاهای املایی می‌توانند به دلایل مختلفی مانند اشتباه در تحریر و یا عدم آگاهی نگارنده از زبان مورد استفاده اتفاق بیفتند. بسیاری از این خطاها یا از دید کاربران پنهان می‌مانند و یا خطاهایی هستند که کاربران به اشتباه گمان بر درست بودن آن‌ها دارند. بروز این خطاهای املایی، کلمات مبهم

بی‌معنی را در پرسش کاربر ایجاد می‌کند که عدم وجود آن‌ها در پایگاه دادگان می‌تواند باعث پاسخگویی نامناسب توسط سیستم پرسش و پاسخ شود. بعلاوه ممکن است کاربر در پرسش خود از لغتی استفاده کند که با وجود املای صحیح، در پایگاه دادگان وجود نداشته باشد. این لغت معنادار نیز با توجه به رویکرد آماری سیستم پرسش و پاسخ طراحی شده، برای سیستم مبهم بوده و می‌تواند باعث پاسخگویی نامناسب سیستم پرسش و پاسخ شود.

در سیستم پرسش و پاسخ طراحی شده، امکان تشخیص و تصحیح خطاهای املایی موجود در پرسش کاربر افزوده شده است. ماژول افزوده شده، قادر است خطاهای املایی را تشخیص دهد و برای اصلاح کلمات نادرست فهرستی از واژگان درست محتمل را پیشنهاد نماید. کاربران می‌توانند از میان فهرست واژگان پیشنهادی، واژه مورد نظر را انتخاب نمایند و یا در صورتی که کلمه مورد نظر آن‌ها در فهرست وجود نداشته باشد، مجددا آن کلمه را وارد نمایند. پس از انتخاب واژه مطلوب توسط کاربر و یا ورود مجدد آن، واژه مذکور به طور خودکار جایگزین واژه نادرست خواهد شد.

در روش پیشنهاد شده جهت تشخیص خطاهای املایی، پرسش کاربر پردازش شده و لغات آن جداگانه بررسی می‌شوند. چنانچه لغتی در پرسش کاربر قید شده باشد، اما در پایگاه دادگان وجود نداشته باشد، یا این لغت املای نادرستی دارد و یا در خوشبینانه‌ترین حالت، لغت مترادف آن در پایگاه دادگان وجود دارد و در هر دو حالت احتمال پاسخگویی صحیح توسط سیستم کاهش می‌یابد.

در روش پیشنهاد شده جهت تصحیح خطاهای املایی، N-گرام‌های مشترک بین لغت مورد نظر و هر یک از لغات پایگاه دادگان مشخص می‌شوند و با توجه به نوع N-گرام‌ها به آن لغت در پایگاه دادگان امتیازی اختصاص می‌یابد (N از یک تا ۴ تغییر می‌کند). برای هر کدام از ۱-گرام‌های مشترک بین دو لغت، امتیاز یک؛ ۲-گرام‌های مشترک، امتیاز دو؛ ۳-گرام‌های مشترک، امتیاز سه و برای هر کدام از ۴-گرام‌های مشترک، امتیاز چهار اختصاص داده می‌شود (افزایش N در N-گرام مشترک، ارزش بیشتری به لغت خواهد داد).

پس از محاسبه امتیاز هر یک از لغات پایگاه دادگان، ۵ لغت با بیشترین امتیاز به کاربر پیشنهاد می‌گردد. کاربر می‌تواند لغت مناسب را انتخاب نموده و یا در صورت تمایل لغت دیگری را تایپ کند. این لغت جایگزین لغت ناشناخته و مبهم شده و پرسش بازسازی می‌شود. سیستم، پرسش بازسازی شده را پردازش نموده و پاسخ مناسب آن را بازمی‌گرداند.

لازم به ذکر است که از ماژول فوق علاوه بر تشخیص و تصحیح کلمات مبهم بی‌معنی (لغات دارای خطای املایی)، جهت تشخیص و جایگزینی کلمات مبهم معنادار (لغاتی که با وجود املای صحیح، در منابع داده وجود ندارند) نیز استفاده می‌شود. دو مثال زیر پرسش‌های دارای خطاهای املایی هستند که مطرح شده‌اند و سیستم به آنها پاسخ داده است.

مثال اول: (ابهام بی معنی)

– کاربر: شرایط صدور معافیت احصیلی برای دانشجویان دوره‌های مجازی چیست؟

– سیستم: لغت احصیلی که در سوال "شرایط صدور معافیت احصیلی برای دانشجویان دوره‌های مجازی چیست" استفاده کردید، در هیچ کدام از متون مرجع وجود نداشت. به جای این لغت یکی از لغات زیر را انتخاب نمایید و یا لغت مورد نظر خود را تایپ کنید.

تحصیلی تحصیلی. فارغ‌التحصیل تحصیلات اصلاحات

کاربر بر روی کلمه "تحصیلی" کلیک می‌کند، کلمه تحصیلی جایگزین کلمه "احصیلی" شده و سیستم پاسخ می‌دهد:

– سیستم: دانش‌آموختگان این دوره‌ها می‌توانند در مقاطع بالاتر ادامه تحصیل دهند و از مزایای استخدامی و اجتماعی همانند سایر دانش‌آموختگان برخوردار گردند. درج عنوان نوع دوره آموزش الکترونیکی در مدارک تحصیلی دانش‌آموختگان الزامی است. صدور معافیت تحصیلی برای دانشجویان نظام یادگیری الکترونیکی مشروط به شرکت در آزمون برگزار شده توسط سازمان سنجش و مصوب بودن مقطع رشته و موسسات توسط شورای گسترش آموزش عالی بلامانع می‌باشد.

مثال دوم: (ابهام معنا دار)

– کاربر: سقف مدت مجاز تحصیل در کارشناسی ارشد مجازی چند سال است؟

– سیستم: لغت سقف که در سوال "سقف مدت مجاز تحصیل در کارشناسی ارشد مجازی چند سال است" استفاده کردید، در هیچ کدام از متون مرجع وجود نداشت. به جای این لغت یکی از لغات زیر را انتخاب نمایید و یا کلمه مورد نظر خود را تایپ کنید.

حقوق تحقق حقوق حقوقی سابقه

– کاربر: "حداکثر" (از آنجا که کلمه "سقف" در مجموعه متون وجود نداشته است، کاربر مترادف آن را تایپ می‌کند)

– سیستم: حداکثر مدت مجاز تحصیل در دوره‌های کارشناسی ارشد ناپیوسته ۲ سال است. درصورتی‌که دانشجو نتواند واحدهای دوره را در این زمان بگذراند از دانشگاه مربوط اخراج می‌شود. در موارد خاص، شورای آموزشی موسسه می‌تواند حداکثر مدت مجاز تحصیل را تا یک سال افزایش دهد.

شکل ۵-۳ نمونه‌ای از تشخیص و تصحیح خطای املایی را نمایش می‌دهد.

شکل ۵-۳ نمونه‌ای از تشخیص و تصحیح خطای املایی

۵-۲-۳ پاسخگویی به سوالات دارای ضمیر

در سیستم پرسش و پاسخ طراحی شده، مجموعه پرسش‌ها و پاسخ‌های ارائه شده به کاربر در حافظه سیستم نگهداری می‌شود تا کاربر بتواند پرسش‌های مرتبط را مطرح نماید. این پرسش‌ها می‌توانند از طریق ضمیر به یکدیگر وابستگی داشته باشند. وجود ضمایر در سوال، دقت پاسخگویی سیستم‌های پرسش و پاسخ را کاهش می‌دهد. از این‌رو پیش از آن که سیستم به سوال کاربر پاسخ دهد، ضمایر موجود در سوال باید با مرجع مناسب جایگزین شوند.

همانگونه که در فصل قبل اشاره شد، سیستم‌های پیشین عمدتا از پارسرهای زبان در استخراج عبارات اسمی و ویژگی‌های نحوی جهت تعیین مرجع استفاده می‌کردند. اما در سیستم طراحی شده از گرامر و معنا استفاده نمی‌شود و تنها با بکارگیری تکنیک‌های آماری، دانش موجود در متون جهت پاسخگویی به سوالات استخراج می‌شود. از این‌رو تعیین و جایگزینی خودکار مرجع ضمایر امکان‌پذیر نیست.

در روش پیشنهاد شده سیستم جهت تعیین مرجع با کاربر تعامل برقرار می‌کند و از این طریق مرجع را جایگزین می‌نماید، به این منظور پس از تشخیص ضمیر در سوال و با در نظر گرفتن این حقیقت که احتمالا مرجع ضمیر در سوال قبل وجود دارد سیستم کلمات پرتکرار و بی معنا را از سوال قبل حذف می‌نماید تا اهمیت کلمات دیگر با توجه به تعداد دفعات تکرارشان در متون تعیین شود. سپس کلمات یافت شده به کاربر پیشنهاد داده می‌شوند. کاربر می‌تواند یکی از این کلمات را به عنوان مرجع انتخاب کند و یا چنانچه مرجع در بین کلمات پیشنهاد شده وجود نداشته باشد، مرجع مورد نظر خود را تایپ نماید. سیستم مرجع را جایگزین ضمیر می‌نماید و پس از بازسازی سوال، پاسخ را از متون بازیابی می‌کند و به کاربر نمایش می‌دهد. دو مثال زیر پرسش‌های دارای ضمایر هستند که مطرح شده‌اند و سیستم به آن‌ها پاسخ داده است.

مثال اول:

- کاربر: نحوه برگزاری امتحانات دوره مجازی چگونه است؟

– سیستم: امتحانات در دوره‌های یادگیری الکترونیکی به صورت حضوری و در محل مرکز آموزش‌های الکترونیکی دانشگاه برگزار می‌گردد.

– کاربر: آیا دانشجو می‌تواند غیبت غیرموجه در آن داشته باشد؟

– سیستم: یکی از عبارات زیر را به عنوان مرجع ضمیر"آن" انتخاب نموده و یا مرجع مناسب را تایپ نمایید؟

برگزاری امتحانات دوره مجازی

کاربر بر روی امتحانات کلیک می‌کند و سیستم سوال اولیه کاربر را به فرم "آیا دانشجو می‌تواند غیبت غیرموجه در امتحانات داشته باشد" بازسازی می‌کند و پاسخ زیر را به کاربر نمایش می‌دهد:

– سیستم: امتحانات در دوره‌های یادگیری الکترونیکی به صورت حضوری و در محل مرکز آموزش‌های الکترونیکی دانشگاه برگزار می‌گردد. غیبت غیرموجه دانشجو در امتحان پایان نیم‌سال هر درس به منزله کسب نمره صفر برای آن درس است. درصورتی‌که غیبت دانشجو در جلسه امتحان پایان نیم‌سال از نظر موسسه موجه تشخیص داده شود درس مزبور حذف می‌شود.

مثال دوم:

– کاربر: آیا امکان ادامه تحصیل برای دانش‌آموختگان دوره‌های مجازی وجود دارد؟

– سیستم: برگزاری دوره کارشناسی ارشد الکترونیکی به شیوه آموزش محور می‌باشد. دانش‌آموختگان این دوره‌ها می‌توانند در مقاطع بالاتر ادامه تحصیل دهند و از مزایای استخدامی و اجتماعی همانند سایر دانش‌آموختگان برخوردار گردند. درج عنوان نوع دوره آموزش الکترونیکی در مدارک تحصیلی دانش‌آموختگان الزامی است.

– کاربر: آیا عنوان آموزش الکترونیکی در مدارک آنها درج می‌شود؟

– سیستم: یکی از عبارات زیر را به عنوان مرجع ضمیر"آنها" انتخاب نموده و یا مرجع مناسب را تایپ نمایید؟

ادامه تحصیل دانش‌آموختگان مجازی

از آنجا که مرجع ضمیر "آنها" عبارت "دانش‌آموختگان دوره‌های مجازی" است و در عبارات پیشنهادی وجود ندارد، کاربر مرجع آن را تایپ می‌کند.

- کاربر: دانش‌آموختگان دوره‌های مجازی
- سیستم: درج عنوان نوع دوره آموزش الکترونیکی در مدارک تحصیلی دانش‌آموختگان الزامی است.

صدور معافیت تحصیلی برای دانشجویان نظام یادگیری الکترونیکی مشروط به شرکت در آزمون برگزار شده توسط سازمان سنجش و مصوب بودن مقطع رشته و موسسات توسط شورای گسترش آموزش عالی بلامانع می‌باشد. آموزش در دوره‌های الکترونیکی در تمام دانشگاه‌های کشور مبتنی بر نظام واحدی است .

شکل ۵-۴ نمونه‌ای از جایگزینی مراجع ضمایر را نشان می‌دهد.

شکل ۵-۴ نمونه‌ای از جایگزینی مراجع ضمایر

۵-۲-۴ پاسخگویی به سوالات حذفی

در سیستم پرسش و پاسخ طراحی شده، مجموعه پرسش‌ها و پاسخ‌های ارائه شده به کاربر در حافظه سیستم نگهداری می‌شود تا کاربر بتواند پرسش‌های مرتبط را مطرح نماید. در برخی موارد، بخشی از پرسش پیرو به قرینه معنایی حذف شده است که آن را وابسته به پرسش قبلی خود می‌نماید. حذف این بخش‌ها دقت پاسخگویی سیستم‌های پرسش و پاسخ را کاهش می‌دهد. از این‌رو پیش از آن‌که سیستم به این‌گونه پرسش‌ها پاسخ دهد، بخش حذف شده آن‌ها باید بازسازی شود.

سوالات حذفی معمولا دارای ساختار کوتاه‌تری می‌باشند. در روش پیشنهاد شده با در نظر گرفتن این حقیقت، سوالاتی که طول آن‌ها ۳ کلمه یا کمتر باشد به عنوان پرسش حذفی شناخته می‌شوند. بخش حذف شده در سوالات حذفی معمولا در سوال قبلی موجود است و به این دلیل سیستم ابتدا کلمات پرتکرار و بی معنا را از سوال قبلی حذف نموده و سپس کلمات کلیدی و با ارزش آن را به سوال فعلی افزوده و جستجو را بر اساس آن انجام می‌دهد. مثال زیر، یک نمونه پرسش حذفی و پاسخ سیستم به آن را نشان می‌دهد.

مثال

– کاربر: آیا امکان انتقال دانشجو در دوره مجازی وجود دارد؟

– سیستم: انتقال به معنی تغییر محل تحصیل دانشجو از یک موسسه به موسسه دیگر در همان رشته و همان مقطع تحصیلی است. انتقال دانشجو در نظام یادگیری الکترونیکی ممنوع می‌باشد. مهمان شدن دانشجویان موسسات داخل و خارج کشور در موسسات و برعکس ممنوع است.

– کاربر: مهمان‌شدن چطور؟

جستجوی سیستم (انتقال دانشجو در دوره مجازی + مهمان‌شدن)

– سیستم: انتقال دانشجو در نظام یادگیری الکترونیکی ممنوع می‌باشد. مهمان‌شدن دانشجویان موسسات داخل و خارج کشور در موسسات و برعکس ممنوع است. دانشجو موظف است حداکثر ظرف

مدت یک ماه پس از دفاع پایان‌نامه نسبت به تسویه حساب با واحدهای مختلف موسسه اقدام نماید.

شکل ۵-۵ نمونه‌ای از پرسش‌های حذفی را نشان می‌دهد.

شکل ۵-۵ نمونه‌ای از پرسش‌های حذفی

۵-۲-۵ بازیابی قطعه‌ای از متون با استفاده از تکنیک‌های آماری

در این تحقیق با توجه به رویکرد آماری موجود از الگوریتمی به نام JIRS جهت تعیین قطعه کوتاهی از متن که شامل پاسخ می‌باشد، استفاده می‌گردد. الگوریتم JIRS یک تکنیک آماری و مستقل از زبان است که بر اساس مدل N-گرام عمل می‌نماید. این الگوریتم به هر قطعه بر اساس تعداد و طول N-گرام‌های مشترک با سوال امتیازی اختصاص می‌دهد و پاراگرافی را که دارای بیشترین امتیاز است به عنوان قطعه‌ای که به احتمال زیاد شامل پاسخ صحیح پرسش کاربر می‌باشد، بازمی‌گرداند.

در این مدل وزن N-گرام‌ها مجموع وزن لغات تشکیل دهنده آن‌ها است و مستقل از مکان لغات می‌باشد. به عنوان مثال وزن N-گرام‌های جملات "the capital of Croatia is" و " is the capital of Croatia" یکسان است. با توجه به اینکه لغات موجود در پرسش‌ها عمدتا همان لغات موجود در متون

هستند و تنها تفاوت، موقعیت آن‌ها در جمله می‌باشد، این الگوریتم از دقت خوبی در بازگرداندن قطعه مناسب برخوردار است [۴۹].

اعمال انجام شده توسط این الگوریتم به ترتیب زیر می‌باشد:

۱ -با استفاده از الگوریتم TF*IDF، به هریک از جملات موجود در متن متناسب با کلمات مشترک آن با پرسش کاربر وزنی اختصاص می‌یابد و سپس از بین آن‌ها جملات با وزن بیشتر بازگردانده می‌شوند.

۲ -از جملاتی که در مرحله قبل انتخاب شده‌اند، برای تشکیل پاراگراف استفاده می‌شود. این پاراگراف از الحاق جملات مجاور به جمله موجود تشکیل می‌شود. به عنوان مثال سند d شامل مجموعه‌ای از n جمله به صورت $d = (s_1, \dots\dots, s_n)$ می‌باشد. برای هر یک از جملات انتخاب شده مانند s_i، پاراگرافی با اندازه m=2k+1 از الحاق جملات $s_{(i-k)}, \dots\dots, s_{(i+k)}$ تشکیل می‌شود. تحقیقات انجام شده توسط گومز [۴۹] نشان داد که با تعیین ۲۰ پاراگراف سه جمله‌ای برای هر پرسش، ۹۰٪ پاسخ‌های بازگردانده شده صحیح می‌باشند. وزن هر یک از لغات موجود در سوال با فرمول زیر محاسبه می‌شود:

$$W_k = 1 - \frac{\ln(n_k)}{1+\ln(N)} \qquad (۱-۵)$$

n_k، تعداد پاراگراف‌های دارای لغت t_k و N تعداد کل پاراگراف‌های مجموعه است. طبق فرمول فوق اگر لغت t_k فقط یک‌بار در پاراگراف ظاهر شود، دارای بیشترین ارزش بوده و وزن آن برابر با یک خواهد بود. همچنین برای کلمات پرتکرار، $n_k = N$ خواهد بود و در نتیجه W_k کمترین مقدار را خواهد داشت.

وزن یک n-گرام، مجموع وزن لغات تشکیل دهنده آن n-گرام است و با فرمول زیر محاسبه می‌شود:

$$h(x) = \sum_{k=1}^{n} W_k \qquad (۲-۵)$$

در فرمول فوق W_1،.... و W_n وزن لغات n-گرام x می‌باشد که از لغات t_1 تا t_n تشکیل شده است:

$$x = t_1\, t_2\, \ldots\ldots t_n \qquad (۳-۵)$$

شباهت پاراگراف p و پرسش q با استفاده از فرمول زیر محاسبه می‌شود:

$$sim(p,q) = \frac{1}{\sum_{i=1}^{n} W_i} \sum_{\forall x \in p} \mathrm{h(x)}\ \frac{1}{d(x,x_{max})} \qquad (۴-۵)$$

Q مجموعه‌ای از n-گرام‌های موجود در پاراگراف p می‌باشد که کلیه لغات آن در پرسش کاربر وجود دارد و P زیرمجموعه‌ای از Q است که شرایط زیر را دارد:

$$p = \{x_1 , \ldots\ldots , x_M\} \qquad (۵-۵)$$

$$\forall x_i \in p: h(x_i) \geq h(x_{i+1})\ \ i \in \{1,2 , \ldots M-1\}$$
$$\forall x, y \in p : x \neq y \rightarrow T(x) \cap T(y) = 0$$
$$min_{x \in p}\ h(x) \geq\ max_{y \in Q-p}\, h(y)$$

در روابط فوق، T(x) مجموعه لغات موجود در n-گرام x است.

ساده‌ترین روش برای اندازه‌گیری فاصله بین دو n-گرام، شمارش تعداد لغات بین آن‌ها است. اما با توجه به کاهش خطی و سریع تابع فاصله فوق با افزایش تعداد لغات، فاصله n-گرام x و n-گرام x_{max} با استفاده از فرمول زیر محاسبه می‌شود:

$$d(x, x_{max}) = 1+\alpha\ \ln(1+L) \qquad (۶-۵)$$

L تعداد لغات بین دو n-گرام، x_{max} ، n-گرام دارای بیشترین وزن و α ثابتی است که با توجه به میزان اهمیت فاصله دو n-گرام مقداردهی می‌شود؛ بهترین مقدار α برابر با ۰٫۱ می‌باشد.

۵-۲-۶ پیشنهاد پرسش‌های پرتکرار

در سیستم‌های پرسش و پاسخ، کاربران عمدتا پرسش‌هایی مطرح می‌نمایند که مرتبط با مجموعه پرسش و پاسخ‌های پیشین می‌باشد و به ندرت موضوع پرسش تغییر می‌نماید. از این ویژگی در سیستم پرسش و پاسخ طراحی شده، استفاده نموده تا پرسش احتمالی بعدی کاربر پیش‌بینی شود. هدف از ارائه این ماژول کمک به حفظ انسجام گفتگو و افزایش دامنه اطلاعات کاربر می‌باشد.

در روش پیشنهاد شده، مجموعه پرسش‌ها و پاسخ‌های ارائه شده در حافظه سیستم نگهداری می‌شوند که با گذشت زمان حجم آن افزایش می‌یابد. سیستم از این مجموعه استفاده نموده، پس از پاسخگویی به پرسش کاربر و با بکارگیری روش N-گرام، میزان شباهت بین پاسخ ارائه شده و هر یک از پرسش‌های پرتکرار را تعیین می‌نماید. سپس ۵ سوال دارای بیشترین شباهت با پاسخ سوال فعلی را به کاربر پیشنهاد می‌نماید. کاربر با کلیک بر روی هر یک از این سوال‌ها می‌تواند آن را انتخاب نموده و یا سوال متفاوت خود را مطرح کند. بدیهی است با انتخاب پرسش از بین مجموعه پرسش‌های پیشنهاد شده، سیستم بدون نیاز به جستجو در متون، پاسخ آن را از مجموعه پرسش و پاسخ‌ها انتخاب نموده و به کاربر ارائه می‌کند. این امر فرایند جستجو را تسهیل و سرعت پاسخگویی را افزایش خواهد داد. شکل ۵-۶ پرسش‌های مرتبط با پاسخ ارائه شده به پرسش "شرایط استفاده از وام دانشجویی چیست؟" را نشان می‌دهد.

شکل ۵-۶ ارائه پرسش‌های مرتبط با پرسش کاربر

۵-۳ ارزیابی سیستم و بررسی نتایج حاصل

ارزیابی سیستم طراحی شده در این تحقیق، با اندازه‌گیری معیارهای کمی و کیفی انجام شده است. به این منظور یک نسخه ساده پرسش و پاسخ ایجاد شده که در آن هیچ تعاملی بین کاربر و سیستم وجود ندارد و با ورود هر پرسش، سیستم با استفاده از تکنیک‌های بازیابی اسناد و بازیابی قطعه متون، پاسخ مناسب را به کاربر ارائه می‌نماید. مقایسه سیستم تعاملی طراحی شده با این سیستم، ملاک ارزیابی کمی و کیفی آن می‌باشد. همچنین از یک مجموعه آموزش و تست به منظور یادگیری و ارزیابی سیستم استفاده شده است که در ادامه معرفی می‌شود.

۵-۳-۱ مجموعه آموزش و تست

در این تحقیق، از سه پایگاه دادگان فارسی با نام WMPR-QA1-2015 ، WMPR-QA2-2015 و WMPR-QA3-2015 جهت آموزش سیستم استفاده شده است (مشخصات سه پایگاه دادگان فوق در بخش ۱-۳ آمده است). همچنین ۸۱ پرسش و پاسخ از مجموعه سه پایگاه دادگان فوق تهیه شده که از آن برای ارزیابی سیستم استفاده شده است. به منظور بررسی کلیه حالت‌های ممکن، این مجموعه شامل ۴۰ پرسش مستقل، ۱۳ پرسش حذفی، ۹ پرسش شامل خطای املایی، ۱۰پرسش شامل ضمیر و ۹ پرسش شامل کلمات ناشناخته می‌باشد.

۵-۳-۲ ارزیابی کمی

جهت ارزیابی کمی، ۸۱ پرسش موجود در مجموعه تست به دو نسخه "استاندارد" و "تعاملی" داده شده است. در این سیستم برای هر پرسش حتما یک پاسخ ارائه می‌گردد، لذا مقادیر به دست آمده توسط معیارهای اف، صحت، دقت و بازخوانی، یکسان می‌باشند (توضیحات بیشتر در بخش ۳-۴-۳ از پایان‌نامه ارائه شده است). همچنین با توجه به اینکه پاسخ‌های ارائه شده به شکل "قطعه

متن" هستند معیار صحت پاسخ، وجود "عبارت اسمی" و یا "هسته اصلی" در قطعه متن بازگردانده شده است. نسخه تعاملی از ۸۱ پرسش مطرح شده، ۷۰ پرسش را به درستی پاسخ داده است و در نتیجه

F-measure=Accuracy=Recall=Precision=70/81=86.4%

این در حالیست که نسخه استاندارد پرسش و پاسخ تنها به ۵۲ پرسش پاسخ صحیح داده و از دقت ٪ ۶۴,۱ برخوردار بوده است.

F-measure=Accuracy=Recall=Precision=52/81=64.1%

با مقایسه مقادیر فوق می‌توان نتیجه گرفت که افزودن تعامل به سیستم‌های پرسش و پاسخ دقت پاسخگویی را ٪ ۲۲,۳ افزایش داده است.

جدول ۵-۱ نمونه‌ای از پرسش‌هایی را که برای آن‌ها سیستم تعاملی پاسخ صحیح و سیستم استاندارد پاسخ اشتباه بازگردانده است، نشان می‌دهد.

جدول ۵-۱ نمونه‌ای از پرسش و پاسخ‌ها در دو سیستم تعاملی و استاندارد

شماره	نوع خطا	پرسش	پاسخ سیستم تعاملی	پاسخ سیستم استاندارد	توضیحات
۱	**کلمه ناشناخته**	اگر معدل کل دانشجو کارشناسی ارشد کمتر از ***چهارده*** شود چه باید کرد؟	میانگین نمرات دروس دانشجو در دوره کارشناسی ارشد در هر نیم‌سال تحصیلی نباید از ۱۴ کمتر باشد اگر نمرات دانشجو در یک نیم‌سال تحصیلی از ۱۴ کمتر باشد دانشجو مشروط قلمداد می‌شود.	حداقل نمره قبولی در هر درس اعم از دروس مصوب دوره یا دروس جبرانی در دوره کارشناسی ارشد ۱۲ است اگر نمره دانشجو در یک درس اختیاری کمتر از ۱۲ باشد می‌تواند با موافقت استاد مربوط به جای آن درس درس دیگری را با پرداخت شهریه آن طبق تعرفه انتخاب نماید.	در پایگاه دادگان به جای چهارده، از ۱۴ استفاده شده و کلمه **چهارده** ناشناخته است.
۲	**کلمه ناشناخته**	کارشناسی ارشد مجازی آموزشی است یا ***پژوهشی***؟	برگزاری دوره کارشناسی ارشد مجازی به شیوه آموزش محور می‌باشد.	دوره‌های آموزشی کارشناسی ارشد مجازی پس از تایید در شورای تحصیلات تکمیلی و شورای دانشگاه درصورتی‌که حداقل ۲۰ نفر دانشجوی ثبت‌نام شده داشته باشند برگزار می‌گردند.	در این پرسش به جای لغت "پژوهش محور" از لغت "پژوهشی" استفاده شده است که در متن وجود ندارد.
۳	**کلمه ناشناخته**	***سقف*** مدت تحصیل در کارشناسی ارشد مجازی چند سال است؟	حداکثر مدت مجاز تحصیل در دوره‌های کارشناسی ارشد ناپیوسته ۲ سال است درصورتی‌که دانشجو نتواند واحدهای دوره را در حداکثر مدت مجاز تحصیل	مدت مجاز استفاده از وام در مدت تحصیل دکترای تخصصی پیوسته حداکثر ۸ سال تحصیلی معادل ۱۶ نیم‌سال	در این پرسش به جای لغت "حداکثر" از لغت "سقف" استفاده شده است که در متن وجود ندارد.

		با موفقیت بگذراند از دانشگاه مربوط اخراج می‌شود.			
۴	**پرسش حذفی**	انتقال چطور؟	انتقال دانشجو در نظام یادگیری الکترونیکی ممنوع می‌باشد.	انتقال اعضای هیئت علمی رسمی قطعی و آزمایشی به موسسات دیگر با موافقت موسسه مبدا و مقصد امکان‌پذیر است.	این پرسش مرتبط، پس از پرسش "آیا امکان مهمان شدن دانشجو در دوره آموزش الکترونیکی وجود دارد" مطرح شده است.
۵	**پرسش حذفی**	حضوری چطور؟	علاوه بر برگزاری کلاس‌های مجازی به صورت هفتگی، یک جلسه کلاس حضوری در هفته اول ترم و یک جلسه کلاس حضوری دیگر دو هفته قبل از برگزاری امتحانات پایان‌ترم دوره کارشناسی ارشد مجازی برگزار می‌شود.	هر واحد درسی در یادگیری الکترونیکی به ازای ۱۵-۱۷ ساعت آموزش حضوری حداقل ۵ ساعت تولید محتوی مفید علمی به صورت الکترونیکی است.	این پرسش مرتبط پس از پرسش "تعداد کلاس غیرحضوری در کارشناسی ارشد مجازی چقدر است" مطرح شده است.
۶	**پرسش حذفی و ضمیر**	تعداد آن چطور؟	تعداد و مبلغ اقساط بهره‌مندان از تسهیلات وام دانشجویی توسط صندوق رفاه دانشجویان تعیین خواهد شد.	شرایط عمومی و آموزشی وام دارا بودن صلاحیت اخلاقی و رعایت شئون دانشجویی عدم استفاده از وام کمک یا بورس تحصیلی اخذ حداقل تعداد واحدهای لازم طبق مقررات آموزشی دانشجوی تمام وقت	این پرسش مرتبط پس از پرسش "مبلغ اقساط وام‌های دانشجویی چگونه تعیین می‌شود" مطرح شده است.

۷	**خطای املایی**	آیا امکان انتقال **بدهی** دانشجو وجود دارد؟	انتقال بدهی دانشجو در یک رشته تحصیلی و یک مقطع در دانشگاه دولتی بلامانع می‌باشد.	انتقال دانشجو در نظام یادگیری الکترونیکی ممنوع می‌باشد.	در این پرسش لغت "یدهی" اشتباها به جای لغت "بدهی" تایپ شده است.
۸	**خطای املایی**	مبلغ **اتساط** وام‌های دانشجویی چگونه تعیین می‌شود؟	تعداد و مبلغ اقساط بهره‌مندان از تسهیلات وام دانشجویی توسط صندوق رفاه دانشجویان تعیین خواهد شد.	دانشجویان انصرافی تارک تحصیل و اخراجی ملزم به پرداخت کامل وام‌های دریافتی و هزینه‌های مربوط به استفاده از خوابگاه دانشجویی می‌باشند.	در این پرسش لغت "اتساط" اشتباها به جای لغت "قسط" تایپ شده است.
۹	**ضمیر**	حداقل چه تعداد دانشجو برای برگزاری این دوره‌ها باید ثبت‌نام کنند؟	دوره‌های آموزشی کارشناسی ارشد مجازی پس از تایید در شورای تحصیلات تکمیلی و شورای دانشگاه درصورتی‌که حداقل ۲۰ نفر دانشجوی ثبت‌نام شده داشته باشند برگزار می‌گردند.	گروه‌های آموزشی بر حسب ضرورت می‌توانند از وجود محققین مذکور برای برگزاری کارگاه‌های آموزشی در زمینه تکنیک‌های جدید علمی، راه‌اندازی آزمایشگاه و کسب خدمات مشاوره با برنامه منظم دعوت نمایند و این دوره‌ها به منزله برنامه ثابت درسی هر ترم در نظرگرفته شود.	ضمیر"این" جایگزین عبارت "دوره مجازی کارشناسی ارشد" شده است.
۱۰	**ضمیر**	انواع آن چیست؟	انواع وام: وام شهریه، وام حج عمره و زیارت ،وام تحصیلی، وام مسکن،وام مسکن متاهلین، وام تحصیلی ممتاز و نمونه	از انواع درآمدهای مذکور در ماده فوق فقط درآمدهای ناشی از عوارض عمومی مشمول پرداخت سهمیه‌ها و حدنصاب‌های مقرر در ماده ۶۸ قانون شهرداری می‌باشد.	ضمیر "آن" جایگزین لغت "وام" شده است.

۵-۳-۳ ارزیابی کیفی

معیار دیگری که برای ارزیابی این سیستم استفاده شده است معیار کیفی می‌باشد. هرچند ارزیابی بسیاری از سیستم‌های پرسش و پاسخ تعاملی، مبتنی بر سیستم بوده و تنها به پاسخ ارائه شده توجه می‌کنند، ارزیابی برخی دیگر از سیستم‌های پرسش و پاسخ تعاملی مبتنی بر کاربر بوده و طراحان میزان رضایت کاربر از کیفیت تعامل را ملاک ارزیابی سیستم خود قرار می‌دهند.

کوارترونی و ماناندهار [۳۷] و کنستانتینوا [۱۸] برای ارزیابی سیستم‌های پرسش و پاسخ خود، پرسشنامه‌ای را در اختیار کاربران سیستم قرار داده‌اند. کاربران پس از پایان فرایند پاسخگویی با دادن امتیاز به هریک از سوالات پرسشنامه، کیفیت تعامل و کارایی سیستم را اعلام می‌نمایند. ارزیابی کیفی سیستم طراحی شده نیز با استفاده از این پرسشنامه که بر اساس استانداردهای PARADISE [۵۰] تهیه شده، انجام شده است (متن پرسشنامه در بخش ۴-۷-۲ ارائه شده است). با این وجود تفاوت در خصوصیات و سطح دانش کاربرانی که از این سیستم‌ها استفاده کرده‌اند مقایسه کیفی سیستم طراحی شده را با آن سیستم‌ها نامعتبر می‌سازد.

در این مرحله، ۵ کاربر آشنا با مفاهیم جستجو و بازیابی اطلاعات انتخاب شدند. کاربران باید پرسش‌های خود را از هر یک از نسخه‌های ساده و تعاملی پرسش و پاسخ مطرح و پس از پایان فرایند پاسخگویی، پرسشنامه زیر را تکمیل می‌کردند:

۱) به چه میزان توانستید اطلاعات مورد نظر خود را از سیستم به دست آورید؟

۲) آیا فکر می‌کنید سیستم در درک منظور شما موفق بوده است؟

۳) آیا دسترسی به اطلاعات موردنظرتان برای شما ساده بوده است؟

۴) آیا درخواست سیستم در بیان مجدد سوال با کلمات یا ساختار جدید به نظرتان منطقی بوده است؟

۵) آیا کار با این سیستم بگونه‌ای بوده است که در آینده مجددا از این سیستم استفاده نمایید؟

۶) آیا سرعت سیستم در پاسخگویی به سوالات شما مناسب بوده است؟

۷) در کل، کیفیت تعامل سیستم با خود را چگونه ارزیابی می‌کنید؟

امتیاز داده شده به هر پرسش، عددی بین یک تا پنج (یک به معنای حداقل امتیاز و پنج به معنای حداکثر امتیاز) است. جدول ۵-۲ میانگین امتیازات وارد شده توسط این کاربران را نشان می‌دهد.

جدول ۵-۲ میانگین امتیازات کاربران

شماره	پرسش	میانگین امتیازات نسخه ساده	میانگین امتیازات نسخه تعاملی
۱	به چه میزان توانستید اطلاعات مورد نظر خود را از سیستم به دست آورید؟	۳,۸	۴
۲	آیا فکر می‌کنید سیستم در درک منظور شما موفق بوده است؟	۳,۲	۳,۷
۳	آیا دسترسی به اطلاعات موردنظرتان برای شما ساده بوده است؟	۳,۴	۳,۹
۴	آیا درخواست سیستم در بیان مجدد سوال با کلمات یا ساختار جدید به نظرتان منطقی بوده است؟	-	۳,۸
۵	آیا کار با این سیستم بگونه‌ای بوده است که در آینده مجددا از این سیستم استفاده نمایید؟	۳,۲	۳,۱
۶	آیا سرعت سیستم در پاسخگویی به سوالات شما مناسب بوده است؟	۲,۹	۲,۶
۷	در کل، میزان رضایت خود از سیستم را اعلام نمایید؟	۳,۷	۴,۱

با مقایسه مقادیر به دست آمده می‌توان موارد زیر را نتیجه‌گیری نمود:

۱) کاربران در استفاده از نسخه تعاملی، اطلاعات بیشتری را کسب نموده‌اند (پرسش ۱).

۲) نسخه تعاملی در درک منظور کاربران موفق‌تر بوده است و این امر می‌تواند به دلیل امکان رفع ابهام توسط این سیستم‌ها می‌باشد (پرسش ۲).

۳) مقادیر به دست آمده نشان می‌دهد کاربران در نسخه تعاملی ساده‌تر به اطلاعات مورد نیاز خود دسترسی پیدا کرده‌اند و این امر می‌تواند به دلیل افزودن ماژول "پیشنهاد پرسش" و

"پیشنهاد عبارات پرتکرار در پرسش" باشد (پرسش ۳).

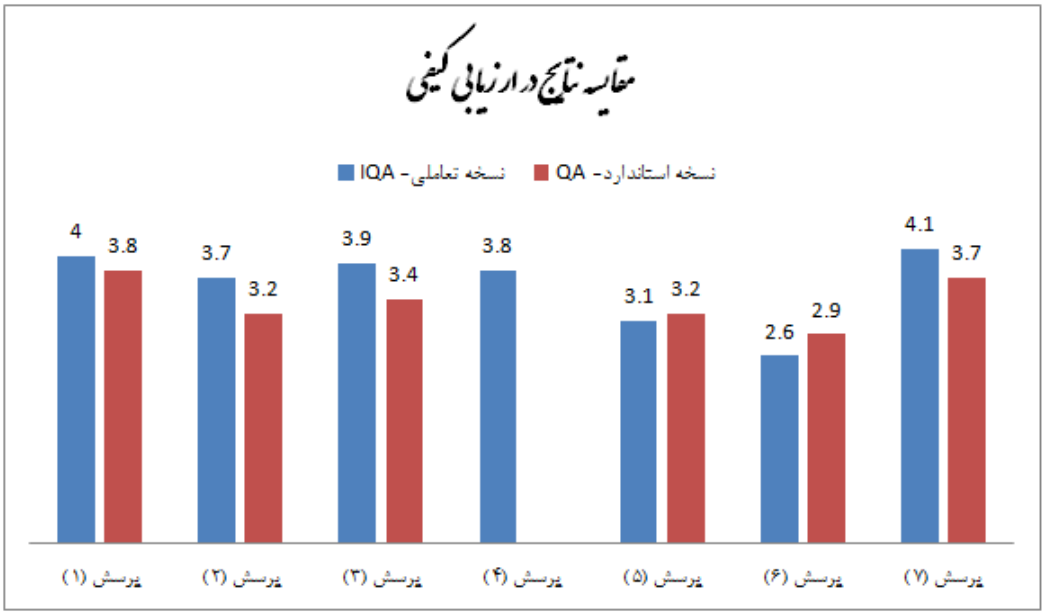
۴) پرسش (۴) فقط جهت ارزیابی سیستم تعاملی طراحی شده است و میانگین به دست آمده، نشان از منطقی بودن پیشنهاد سیستم در بازسازی پرسش‌ها دارد.

۵) متاسفانه کاربران تمایل بیشتری به استفاده مجدد از سیستم پرسش و پاسخ استاندارد اظهار داشته‌اند و در برخی موارد نیز این‌گونه توضیح داده‌اند که "در صورت افزایش سرعت پاسخگویی از سیستم تعاملی استفاده خواهند نمود" (پرسش ۵).

۶) مقادیر به دست آمده برای پرسش (۶)، نشان می‌دهد که سرعت پاسخگویی در سیستم پرسش و پاسخ استاندارد بیشتر از نسخه تعاملی می‌باشد و دلیل این امر می‌تواند انجام برخی پردازش‌ها در نسخه تعاملی در سمت سرور باشد.

۷) کاربران از هر دو سیستم رضایت نسبی داشته‌اند، البته میزان رضایت در سیستم پرسش و پاسخ تعاملی کمی بیشتر بوده است (پرسش ۷).

لازم به ذکر است که میانگین امتیازات در هر پرسش بیشتر از ۲,۵ بوده که بیانگر رضایت نسبی کاربر در آن پرسش است. شکل ۵-۷ نمودار ارزیابی کیفی سیستم طراحی شده را نمایش می‌دهد.



شکل ۵-۷ نمودار ارزیابی کیفی سیستم طراحی شده

فصل ششم

جمع بندی و کارهای آینده

نظام پرسش و پاسخ اتوماتیک، زیرشاخه‌ای از علم بازیابی اطلاعات است که در آن کاربر پرسش خود را به زبان طبیعی مطرح می‌کند و سیستم پاسخ کوتاهی را که احتمال می‌دهد مناسب‌ترین پاسخ باشد به وی بازمی‌گرداند. با توجه به ابهامات موجود در زبان طبیعی، تلاش برای فهم پرسش کاربر و بازگرداندن پاسخ مناسب همواره مورد توجه پژوهشگران این حوزه بوده است. در این تحقیق سعی شده است با بکارگیری تکنیک‌های آماری و بدون توجه به روابط معنایی و گرامر، یک سیستم پرسش و پاسخ مستقل از زبان طراحی شود. در سیستم طراحی شده از دانش نهفته در متون ساختارنیافته برای ایجاد تعامل با کاربر و رفع ابهامات، جهت افزایش دقت پاسخگویی استفاده شده است. همچنین این سیستم با ارائه پیشنهادات مناسب با توجه با تاریخچه گفتگوها به کاربر کمک می‌کند تا پرسش خود را بهتر مطرح کند و در نتیجه پاسخ مناسب‌تری از سیستم دریافت نماید. یکی دیگر از نتایج این تحقیق، استخراج و تهیه اولین پایگاه دادگان فارسی برای طراحی و ارزیابی سیستم‌های پرسش و پاسخ فارسی می‌باشد. در این تحقیق سه پایگاه دادگان فارسی با نام‌های WMPR-QA1-2015، WMPR-QA2-2015 و WMPR-QA3-2015 تهیه شده است که هم اکنون از آزمایشگاه وب‌کاوی و شناسایی الگو دانشگاه صنعتی شاهرود[1] قابل دریافت می‌باشند.

به عنوان ایده‌های پژوهشی برای ادامه کار در حوزه سیستم‌های پرسش و پاسخ تعاملی می‌توان به موارد زیر اشاره نمود:

۱) یافتن راهکارهایی به منظور ترکیب و خلاصه سازی قطعه متون به دست آمده از اسناد مختلف، می‌تواند یکی از حوزه‌های پژوهشی باشد. انجام این پژوهش می‌تواند به درک بهتر چگونگی ادغام چند پاسخ توسط سیستم و ارائه نتیجه به کاربر در قالب یک پاسخ واحد کمک نماید.

۲) در این تحقیق، مرتبط بودن یک پرسش با پرسش قبل از آن بررسی شده است. به عنوان مثال در صورت وجود ضمیر در یک پرسش، سیستم طراحی شده مرجع آن را در پرسش قبل جستجو می‌نماید. باید توجه داشت که سوالات می‌توانند به صورت مجموعه‌ای مرتبط با هم مطرح شوند و

[1] http://wmpr.ir/fa/index/category/53

بنابراین مرجع یک ضمیر ممکن است در پاسخ قبل و یا حتی در تاریخچه گفتگوی کاربر و سیستم موجود باشد. از این‌رو ارائه راهکاری جهت بررسی تاریخچه گفتگوها ضروری است. بررسی زنجیره مارکوف به عنوان یک راهکار مناسب، می‌تواند یکی از اهداف پژوهشی آینده محسوب شود.

۳) از آنجا که هدف این تحقیق، بازیابی اطلاعات از متون ساختارنیافته می‌باشد، باید امکان وجود ابهامات در متون و راهکار مناسب جهت رفع آن‌ها بررسی گردد. به عنوان مثال منابع داده ممکن است شامل ضمایر باشد که تعیین و جایگزینی مراجع آن‌ها، دقت پاسخ را افزایش خواهد داد.

پیوست

الف) فرم ارزیابی کیفی

بسمه تعالی

کاربر عزیز:

با سلام و احترام خواهشمند است پرسشنامه زیر را که جهت ارزیابی عملکرد سیستم مورد استفاده شما طراحی شده است، مطالعه فرموده و به سوالات آن با دادن امتیاز مورد نظرتان از یک تا پنج پاسخ دهید. از جنابعالی که با دقت نظر و صرف وقت ما را در ارزیابی سیستم یاری می‌فرمایید کمال تشکر را دارم.

ردیف	**لطفا نظر خودتان را در هر یک از موضوعات زیر با عددی بین ۱ تا ۵ بیان نمایید:**	**امتیاز در سیستم** QA	**امتیاز در سیستم** IQA	**توضیحات**
۱	به چه میزان توانستید اطلاعات مورد نظر خود را از سیستم به دست آورید؟			
۲	آیا فکر می‌کنید سیستم در درک منظور شما موفق بوده است؟			
۳	آیا دسترسی به اطلاعات موردنظرتان برای شما ساده بوده است؟			
۴	آیا درخواست سیستم در بیان مجدد سوال با کلمات یا ساختار جدید به نظرتان منطقی بوده است؟			
۵	آیا کار با این سیستم بگونه‌ای بوده است که در آینده مجددا از این سیستم استفاده نمایید؟			
۶	آیا سرعت سیستم در پاسخگویی به سوالات شما مناسب بوده است؟			
۷	در کل، کیفیت تعامل سیستم با خود را چگونه ارزیابی می‌کنید؟			

فهرست مراجع

[1] Kolomiyets O., Moens M.F., (2011), "A survey on question answering technology from an information retrieval perspective," *Information science, 181(24),* pp. *5412-5434.*

[2] Green B.F., et al., (1961), "Baseball: an automatic question-answerer," Proceedings Western Computing Conference, 19, pp. 219-224, New York.

[3] Voorhees E.M., Tice D.M., (2000), "Building a question answering test collection," Proceedings 23rd annual international ACM SIGIR conference on Research and development in information retrieval, pp. 200-207, Greece.

[4] آفتابی ز، (۱۳۹۳)، گزارش سمینار: "مروری بر سیستمهای پرسش و پاسخ مبتنی بر بازیابی از اطلاعات"، دانشکده مهندسی برق و کامپیوتر، دانشگاه یزد

[5] Gobeill J., et al., (2009), "Question Answering for Biology and Medicine," Proceedings 9th International Conference on Information Technology and Applications in Biomedicine (ITAB 2009), pp. 1-5, Cyprus.

[6] Radev D., Qi H., Wu H., Fan W., (2002), "Evaluating Web-based Question Answering Systems," Proceedings the 3th International Conference on language resources and evaluations (LREC 2002) , Las Palmas.

[7] Roberts I., Gaizauskas R., (2004), "Evaluating Passage Retrieval Approaches for Question Answering," Proceedings 26rd European Conference on Information Retrieval, pp. 72-84, Sunderland

[8] Katz B., Borchardt G., Felshin S., (2006), "Natural Language Annotations for Question Answering," Proceedings 19th International FLAIRS Conference (FLAIRS 2006), Melbourne Beach.

[9] Katz B., (1997), "Annotating the World Wide Web Using Natural Language," Proceedings 5th RIAO Conference on Computer Assisted Information Searching on the Internet (RIAO '97), pp. 136-155, Montreal.

[10] Katz B., et al., (2004), "Viewing the web as a virtual database for question answering," pp. 215-226, In: "*New Directions in Question Answering*", ISBN 0-262-63304-3, Maybury M.T., MIT Press, Cambridge.

[11] Weizenbaum J., (1966), "Eliza - a computer program for the study of natural language communication between man and machine," *Communications of the*

ACM, 9(1), pp. *36-45*.

[12] Bobrow D., Kaplan R., Kay M., Norman D., Thompson H., Winograd T., (1997), "Gus: a frame-driven dialog system," *Artificial Intelligence, 8(2),* pp. *155-173*.

[13] Qu Y., Green N., (2002), "A constraint-based approach for cooperative information seeking dialogue," Proceedings of the Second International Natural Language Generation Conference ACL.

[14] Rieser V., Lemon O., (2009), "Does this list contain what you were searching for learning adaptive dialogue strategies for interactive question answering," *Natural Language Engineering, 15(1),* pp. *55-72*.

[15] Dornescu I., Orasan C., (2010), "Interactive qa using the qall-me framework," *International Journal of Computational Linguistics and Applications, 1(1-2),* pp *233-247*.

[16] Quarteroni S., Manandhar S., (2009), "Designing an interactive open-domain question answering system," *Natural Language Engineering, 15(1),* pp *73-95*.

[17] Tang Y., Zheng Z., (2011), "Towards Interactive QA: suggesting refinement for Questions," SIGIR Workshop on entertain me, pp. 13-14, China.

[18] Konstantinova N., Orasan C., (2013), "Interactive question answering," pp. 149-169, In: "*Emerging Applications of Natural Language Processing: Concepts and New Research*", S. Bandyopadhyay, S.K. Naskar, A. Ekbal, IGI Global.

[19] Laokulrat N., (2013), "A survey on question classification techniques for question answering," *KMITL Information Technology Journal, 2(1).*

[20] Dwivedi S.K., Singh V.¸ (2013), "Research and Reviews in Question Answering System," Proceedings International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA), 10, pp. 417-424, India.

[21] Yen S.J., et al., (2013), "A support vector machine-based context-ranking model for question answering," *Information science, 224,* pp. *77-87*.

[22] Lashkari A.H., Mahdavi F., Ghomi V., (2009), "A Boolean Model in Information Retrieval for Search Engines," Proceedings International Conference on Information Management and Engineering (ICIME '09), pp. 385-389, Malaysia.

[23] Brill E., Dumais S., Banko M., (2002), "An Analysis of the AskMSR Question Answering System," Proceedings conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 257-264, Philadelphia.

[24] Carvalho G., de Matos D.M., Rocio V., (2007), "Document retrieval for question

answering: a quantitative evaluation of text preprocessing," Proceedings the ACM first Ph.D. workshop in CIKM (PIKM '07), pp. 125-130, New York.

[25] Mishra M., Kumar V., Sharma H.R., (2013), "Question Classification using Semantic, Syntactic and Lexical features," *International Journal of Web & Semantic Technology (IJWesT), 4(3),* pp. *39-47*.

[26] Habernal I., Konopik M., Rohlik O., (2012), "Question answering," pp. 304-343, In: "*Next Generation Search Engines: Advanced Models for Information Retrieval*", C. Jouis, I. Biskri, M. Roux, IGI Global.

[27] Moldovan D., Harabagiu S., Pasca M., Mihalcea R., Girju R., Goodrum R., Rus V., (2000), "The structure and performance of an open-domain question answering system," Proceedings the Conference of the Association for Computational Linguistics (ACL-2000), pp. 563-570, Hong Kong.

[28] Harabagiu S., Moldovan D., (2003), "Question answering," pp. 560-582, In: "*The Oxford Handbook of Computational Linguistics*", R. Mitkov, Oxford University Press, New York.

[29] Webber W.E, (2010), Ph.D Thesis, "Measurment in Information Retrieval Evaluation," The university of Melburne.

[30] میدو چارلزتی، کرافت دونالد اچ، بویس برت‌آر، باری کارول، (۱۳۹۰)، "نظامهای بازیابی اطلاعات متنی"، ترجمه نجلا حریری، انتشارات چاپار، تهران

[31] حریری ن، بابالحوائجی ف، فرزندی‌پور م، نادی راوندی س، (۱۳۹۲)، "معیارهای ارزیابی ربط در نظامهای بازیابی اطلاعات: دانسته‌ها و ندانسته‌ها"، پژوهشنامه پردازش و مدیریت اطلاعات (۱)۳۰، ص ۱۹۹-۲۲۱

[32] Schutze H., (2013), "Introduction to Information Retrieval," Relevance Feedback and Quary Extention, pp. 205-209, London.

[33] Nan Ye., Chai K.M., Lee W.S, Chieu H.L, (2012), "Optimising F-Measures: A Table of Tow Approaches," Proceeding of the 29th Intern ational Conference on Machine Learning, pp. 198-204, Edinburgh.

[34] Manning Ch., Raghavan P., Schütze H., (2008), " Introduction to Information Retrieval, Cahpter 1: Evaluation in Information Retrieval," Cambridge University Press, Cambridge, pp. 152-162.

[35] Forner P., Giampiccolo D., Magnini B., Peñas A., Rodrigo Á., Sutcliffe R., (2010), "Evaluating multilingual question answering systems at CLEF", Proceedings of the Seventh International Conference on Language Resources and Evaluation

(LREC'10), pp. 2774-2781, Malta.

[36] Yao X., (2014), PhD. Thesis, "Feature-driven Question Answering with Natural Language Alignment", Department of Computer Science, Johns Hopkins University.

[37] Cheongjae L., Sangkeun J., Kyungduk K., Donghyeon L., Geunbae L., (2010), "Recent Approaches to Dialog Management for Spoken Dialog Systems," *Journal of Computing Science and Engineering, 4(1)*, pp. *1-22*, Korea.

[38] Quarteroni S., (2007), PhD. Thesis, "Advanced techniques for personalized, interactive question answering," Department of Computer Science, The University of York, York, UK.

[39] Smith J., (2010), Master's thesis, "Iqabot: A chatbot-based interactive question-answering system," York University.

[40] Fukumoto J., (2006), "Answering questions of Information Access Dialogue (IAD) task using ellipsis handling of follow-up questions," Proceedings Workshop on Interactive Question Answering (HLTNAACL 06), pp. 262–295, New York.

[41] Mingyu S., Joyce Y., (2007), "Discourse Processing for Context Question Answering Based on Linguistic Knowledge," Know.-Based Syst, 20(6), pp.511–526.

[42] Wanzare L., (2010), Technical Report, "Coreference Resolution," Computational Linguistics Department, Saarland University, Germany.

[43] Soon M.W., Ng T.H., Lim Y.D., "A Machine Learning Approach to Coreference Resolusion of Noun Phrase," *Computational Linguistics, 4(27)*

[44] Moosavi N., Ghassem-Sani GH., (2009), "A Ranking Approach to Persian Pronoun Resolution," *Advances in Computational Linguistics. Research in Computing Science, 41*, pp. 169-180,

[45] Fallahi F., Shamsfard M., "Recognizing Anaphora Reference in Persian Sentences," IJCSI, 2010

[46] د. نوگورانی، م. صبوریان، (۱۳۸۴)،"طراحیو پیاده‌سازی خطایاب فارسی"، دومین همایش فارسی و رایانه، تهران.

[47] م. رسولی، ب. مینایی بیدگلی، (۱۳۸۷)، "روشی جدید در خطایابی املایی در زبان فارسی"، دومین کنفرانس داده کاوی ایران، تهران

[48] Harabagiu S., Hickl A., Lehmann J., Moldovan D., (2005), "Experiments with interactive question-answering," Proceedings of the 43rd Annual Meeting on

Association for Computational Linguistics, pp. 205-214, USA.

[49] Buscaldi D., Rosso P., Sanchis E., (2009), "Answering questions with an n-gram based passage retrieval engine," *Springer Science*, DOI *10.1007/s10844-009-0082-y*.

[50] Walker M.A., Litman D.J., Kamm C.A., Abella A., (1998), "Evaluating spoken dialogue agents with PARADISE: Two case studies," *Computer Speech and Language*, *12*, pp. *317-347*

Abstract. In question answering systems, users ask their questions in a natural language and the system returns the one or more possible answers. However, in question answering systems there is no continued dialogue between the user and the system and the users cannot explain their meaning if it is misunderstood by the system. Hence, getting access to the needed information is not so easy for the user. Adding interaction to these systems enables users to ask related questions and to clarify their meaning and increases the accuracy of the answers returned by the system.

The purpose of this research is to design an interactive question answering system which uses statistical techniques to extract the needed information from unstructured texts. In contrast to other systems which make use of linguistic techniques, this system does not use grammar and semantic techniques and so, it is independent of language. It can answer the questions asked in any language if the appropriate dataset in that language is given to it. In this research, the designed system makes use of a Persian dataset for training and testing. The efficiency of the designed system was assessed by both qualitative and quantitative measures. The obtained scores indicate a 22.3% increase in the accuracy of the answers given by the designed system in comparision with the answers given by the QA system. Moreover, the examination of the users' comments, clearly showed their satisfaction of their interaction with the designed system.

Keywords: Interactive Question Answering System, Information Retrieval, Question Answering System, Human and Computer Interaction

E-learning Center, University of Shahrood
Faculty Computer Engineering

An Interactive Question Answering System Using Artificial Intelligence Techniques

Salime Sadat Shahraini

Supervisor: Dr. Morteza Zahedi
Co-Supervisor: Mehdi Yaghobi, M.Sc.

Date: Auguest 2015