

صلى الله عليه وسلم

دانشگاه یزد

دانشکده مهندسی برق و کامپیوتر

گروه مهندسی کامپیوتر

پایان نامه

برای دریافت درجه کارشناسی ارشد

مهندسی کامپیوتر - هوش مصنوعی-رباتیک

ارائه راهکاری برای تحلیل پاسخ نهایی در
سیستم‌های پرسش و پاسخ برای پرسش‌های با
پاسخ کوتاه

استاد راهنما

دکتر محمدرضا زارع میرک آباد

استاد مشاور

مهندس حسن بشیری

پژوهش و نگارش

صفورا طلیمیان

مهر ماه ۱۳۹۳

به نام، یاد و امیدش

هزار سال رشک می بردم بر پدرم آدم و مادرم حوا که حضورت را لمس می کردند و تورا به نام می خواندند و پا سخت را به
کوش سمری شنیدند و در بهشت خانه داشتند، اما امروز که گردش چرخ مینایی به غربتم کشید ایمان یافتیم که هزاران بار از پدرم آدم و
مادرم حوا به درگاهت نزدیک ترم. چرا که آنان لمس نکردند آغوش مادری، نشنیدند استواری صدای پدر را و آرام نگر فتند در هرم
نفس های این دو فرشته؛ که در آیت روشن بخشش توست و مادر آینه تمام نمای مهرت.

صد سگر که امنیت صدای پدر و آرامش حضور مادر هزار بهشت در این دنیای دون برایم ساخت و چه بی منت تمام

بهشتیان آرا نگاه حضورم شد.

به پاس تنها یک اشک چشم مادر و یک لحظه حضور گرم پدرم این رساله که هر چند برگ سبز است ولی رویداده در اشک چشم
و بارور به خون دل با هزار شرم حضور تقدیم دستان گرم و نگاه مهربانشان می کنم که نه جبران لحظه ای حضورشان بلکه عامل
رویدن گل لبخندی حتی برای کم تر از ثانیه ای بر صورت چون خورشیدشان باشد.

تدیرونگ

سپاس آن بی‌همال و یار و با قدرت توانارا کزو یابد توانایی به عالم هر توانارا

ناصر خسرو

سپاس مخصوص ذات آن پاک بی‌همتاست که کل هستی بدون نگاه مکارنده او لا وجود و بی‌اثرند. و چه خوش مکارنده ایست و چه بی‌فروغ

است چشم مادر دیدن جلوه عشق او. چنان که عطار حق سخن این چنین ادا کرد «کرد از جان مرد موسیقی شناس سخن موسیقی هستی را

سپاس».

اوست مبداء علم و تمام عالم نقطه پرگار اویند و حکم همان که او فرمود. که به پاس رحمت و عشق در طریق دانش چراغ هدایت برافروخت
که غریقان دریای علم حلاک طوفان تردید نگردند، تا این چنین آموختن را سراغازی زیبا نهند.

شعورترین سخن در اندیشه ام حین مکارش این سطور کلام «من لم یسکر المخلوق ولم یسکر الخالق» بود. لذا به رسم ادب و حرمت
آموختن صمیمانه ترین تقدیر را در خالصانه ترین لغات به جناب آقای دکتر محمد رضا زارع میرک آباد و جناب آقای مهندس حسن بشیری تقدیم
می‌کنم که اندیشیدن را پیش از بکارگیری علم و تلاش را پیش از توانایی با اندیشه ام پیوند زدند.

صدافوس که لغات قاصرند از سپاس نگاه های نگران و قلب های عاشق که اگر چنین بود شهد شیرین تمام لغات را از سطر سطر این
کاغذ به کام پدر و مادری می‌ریختم که حضورشان با هیچ کنجی قابل قیاس نیست و افوس که صبر شرمنده بهمنامش شد و مراد یون صوری خواهی
کرد که آرامش محطه هایم مرهون حضور اوست.

باشد که بودن شمرنش امیدشان باشد.

چکیده

با توجه به حجم عظیم اطلاعات پراکنده در پهنه وب و عدم توانایی سیستم‌های بازیابی اطلاعات در استخراج پاسخ مختصر و جامع به کاربر، سیستم‌های پرسش و پاسخ مطرح شدند. این سیستم‌ها در ابتدا تنها به حضور یا عدم حضور لغات برای استخراج پاسخ توجه می‌کردند. اما در توسعه‌های بعدی نقش نحوی و معنایی لغات نیز در نظر گرفته شد. از طرف دیگر ذهن بشر، به عنوان مهم‌ترین الگو، در استخراج پاسخ مبتنی بر تفکر شی گرای می‌کند.

در این پژوهش با استفاده از روابط مفهومی بین لغات و با استناد به دو رابطه `has-operation` و `has-attribute` ایده شی‌گرایی در دنیای واقعی مدل می‌شود. با استفاده از این ایده می‌توان تفکر انسانی را در پاسخ‌گویی به سوالات شبیه‌سازی کرد. از سوی دیگر انسان در پاسخ به سوالات ابتدا به دنبال تعریف مفاهیم می‌گردد. بر این اساس، سوالات پاسخ کوتاه به چند سوال تعریفی تبدیل شده و پاسخ آنها از بین این تعاریف استخراج می‌شود. بنابراین در روش پیشنهادی ابتدا مفاهیم و روابط به سیستم ارائه شده و سیستم با استفاده از این روابط به دنبال پاسخ سوال می‌گردد. نهایتاً سیستم، پس از ارائه پاسخ و دریافت بازخورد از کاربر، به تحلیل روابط مفهومی پرداخته و بر اساس امتیاز روابط، روابط پر اهمیت را شناسایی می‌کند.

روش پیشنهادی روی متون دریافتی از موتورهای جستجو در حوزه‌ی فناوری اطلاعات و ارتباطات و با داده‌های واقعی تست شد. نتیجه این بررسی‌ها بیانگر توانایی روابط معنایی مبتنی بر شی‌گرایی در مدل‌سازی مفاهیم است. این روش به دلیل تکیه بر اشیاء، به عنوان بخش اصلی سخن، و تعریف رفتار، به عنوان وابستگان به اشیاء، در مواجهه با سوالات فاقد فعل بخوبی عمل می‌کند.

کلمات کلیدی: سوالات پاسخ کوتاه، سیستم‌های پرسش و پاسخ، روابط معنایی، داده‌های

متنی

فهرست مطالب

۱	فصل اول
۵	۱-۱ اهمیت موضوع
۶	۱-۲ فرضیات و هدف
۶	۱-۳ مراحل تحقیق و مشکلات
۷	۱-۴ ساختار پایان نامه
۹	فصل دوم
۱۱	۲-۱ مفاهیم اولیه
۱۴	۲-۲ معماری سیستم‌های پرسش و پاسخ
۱۶	۲-۲-۱ فاز اول پردازش سؤال
۱۸	۲-۲-۲ فاز دوم بازیابی اطلاعات
۱۹	۲-۲-۳ فاز سوم انتخاب بهترین پاسخ
۲۰	۲-۳ معرفی راهکارهای پاسخ به سؤال در سیستم‌های پرسش و پاسخ
۲۳	۲-۳-۱ روش نمایش به صورت کیسه‌ای از کلمات
۲۶	۲-۳-۲ تجزیه و تحلیل ریخت‌شناسی - زبانی اجزای زبان طبیعی
۲۹	۲-۳-۳ طبقه‌بندی معنایی پاسخ مورد انتظار
۳۳	۲-۳-۴ طبقه‌بندی تمام اجزا سؤال و یا ساختار زبان طبیعی
۳۶	۲-۳-۵ شناخت روابط گفتمان لازم
۳۸	۲-۳-۶ ترجمه و بازیابی با یک زبان ساخت یافته
۳۹	۲-۳-۷ ترجمه و بازیابی با یک نمایش منطقی
۴۳	فصل سوم
۴۶	۳-۱ معرفی بستگی‌های استنفورد
۴۷	۳-۲ روابط معنایی وردنت ۳

۴۸.....	روابط عام/خاص	۳-۲-۱
۴۸.....	روابط کل / جزء	۲-۲-۳
۴۹.....	رابطه استلزام	۳-۲-۳
۵۰.....	رابطه علیت	۴-۲-۳
۵۰.....	رابطه شباهت	۵-۲-۳
۵۰.....	رابطه همچنین	۶-۲-۳
۵۱.....	رابطه خصلت	۷-۲-۳
۵۱.....	گروه‌بندی افعال	۸-۲-۳
۵۱.....	دامنه و عضو دامنه	۹-۲-۳
۵۲.....	روابط تکمیلی وردنت	۳-۳-۳
۵۳.....	has_at t r i but e	۱-۳-۳
۵۳.....	has_ope r at i on	۲-۳-۳
۵۳.....	Causes / Causes By	۳-۳-۳
۵۳.....	Pr ecedes / Comes af t er	۳-۳-۴
۵۴.....	Hbl o_event / Mer o_event	۳-۳-۵
۵۴.....	I s_ used_ f or	۶-۳-۳
۵۴.....	نحوه استخراج روابط تکمیلی وردنت	۴-۳-۳
۵۷.....	فصل چهارم	
۵۹.....	۱- شیوه درک مفاهیم	۴-۴-۱
۶۰.....	۲- نحوه پاسخ‌گویی به سؤال توسط انسان	۴-۴-۲
۶۳.....	۳- ارائه راهکاری برای پاسخ به سؤالات پاسخ کوتاه	۴-۴-۳
۶۴.....	۴- معماری روش پیشنهادی	۴-۴-۴
۶۴.....	۱- فاز اول پردازش سؤال	۴-۴-۱

۶۵	۴-۴-۱-۱	روش پاسخ‌گویی به سؤالات تعریفی
۶۵	۴-۴-۱-۲	روش پاسخ‌گویی به سؤالات پاسخ کوتاه
۶۶	۴-۴-۲	معرفی پایگاه روابط موجود
۶۸	۴-۴-۳	فاز دوم بازیابی اطلاعات
۶۸	۴-۴-۴	فاز سوم استخراج پاسخ
۶۹	۴-۴-۵	فاز چهارم ارائه پاسخ و دریافت بازخورد
۷۰	۴-۴-۶	فاز پنجم تحلیل رابطه
۷۱	۴-۵	مثال عملکرد روش
۷۵		فصل پنجم
۷۷	۵-۱	مجموعه روابط
۷۷	۵-۲	پیاده‌سازی فاز اول
۷۸	۵-۳	تشخیص سؤالات تعریفی معادل سؤال پاسخ کوتاه
۷۹	۵-۴	ارسال به موتور جستجو
۷۹	۵-۵	تجزیه اسناد
۸۰	۵-۶	استخراج روابط معنایی و امتیازدهی به پاراگرافها
۸۰	۵-۷	رتبه‌بندی پاراگرافها و ارائه بهترین پاسخ
۸۱	۵-۸	دریافت بازخورد از کاربر و بروز رسانی روابط
۸۲	۵-۹	استخراج رابطه <code>has_attribute</code> و <code>has_operation</code>
۸۵		فصل ششم
۸۷	۶-۱	راهکارهای ارزیابی سیستم پرسش و پاسخ
۹۰	۶-۲	مجموعه سؤالات برای تست سیستم
۹۰	۶-۳	روش‌های قابل مقایسه با راهکار پیشنهادی
۹۰	۶-۴	نتایج

۹۱.....	۱-۴-۶ بررسی عملکرد سیستم.....
۹۲.....	۲-۴-۶ تأثیر روابط معنایی has_at t r i but e و has_ o p e r a t i o n
۹۴.....	۳-۴-۶ مقایسه راهکار ارائه شده با روش طبقه بندی تمام اجزاء سؤال.....
۹۶.....	۴-۴-۶ مقایسه راهکار ارائه شده با روش تجزیه و تحلیل ریخت شناسی-زبانی.....
۹۸.....	۵-۴-۶ نتیجه گیری و کارهای آینده.....
۸۷.....	۷ پیوست
۹۲.....	۸ منابع:

فهرست جدول‌ها

- جدول ۱-۲ : گروه‌بندی ارائه‌شده توسط جان سوزوکی در سال ۲۰۰۳ [۲۵]..... ۳۱
- جدول ۲-۲ : بخشی از دسته‌بندی ارائه‌شده با استفاده از طبقه‌بندی دولایه [۲۶]..... ۳۲
- جدول ۱-۳ : برخی از بستگی‌های استنفورد [21]..... ۴۷
- جدول ۱-۴ : روابط معنایی مورد استفاده در راهکار پیشنهادی..... ۶۷
- جدول ۲-۴ : روابط معنایی کلمه array در مجموعه روابط سیستم..... ۷۲
- جدول ۱-۶ : امتیاز Pyramid-score در ۳۵ اجرا..... ۹۱
- جدول ۲-۶ : تأثیر دو رابطه has_operation و has_attribute بر کارایی سیستم..... ۹۳
- جدول ۳-۶ : مقایسه راهکار پیشنهادی با روش طبقه‌بندی اجزا سؤال..... ۹۵
- جدول ۴-۶ : محاسبه دقت کلی در سیستم پیشنهادی..... ۹۷
- جدول ۵-۶ : مقایسه روش پیشنهادی با روش تجزیه و تحلیل ریخت‌شناسی زبانی..... ۹۷

فهرست شکل‌ها

شکل ۱-۲ : نمایش عملکرد کلی سیستم‌های پرسش و پاسخ **Error! Bookmark not**

defined.

شکل ۲-۲ : معماری سیستم پرسش و پاسخ متنی [۸].....۱۵

شکل ۳-۲ : معماری کلی سیستم پرسش و پاسخ Ephyra [۱] **Error! Bookmark not**

defined.

شکل ۴-۲ : نمایش بخش پردازش سؤال در سیستم‌های پرسش و پاسخ [۱].....۱۷

شکل ۵-۲ : واحد بازیابی اطلاعات در سیستم‌های پرسش و پاسخ [۱].....۱۸

شکل ۶-۲ : واحد انتخاب پاسخ در سیستم پرسش و پاسخ [۱].....۱۹

شکل ۷-۲ : تطبیق نیاز اطلاعاتی و اسناد در روش نمایش کیسه‌ای از کلمات [۱].....۲۴

شکل ۸-۲ : نمونه‌ای از گراف وابستگی بر اساس روابط استنفورد [۲۲].....۲۹

شکل ۹-۲ : قالب ارائه شده برای فعل killing [۳۰].....۳۴

شکل ۱-۳ : نحوه استخراج رابطه has_operation [۶۰].....۵۵

شکل ۱-۴ : بخشی از شبکه لغت binary tree.....۶۰

شکل ۲-۴ : روند پاسخ به سؤال توسط انسان.....۶۱

شکل ۳-۴ : نمایش معماری سیستم پیشنهادی.....۶۴

شکل ۴-۴ : واحد پردازش سؤال در راهکار پیشنهادی.....۶۶

شکل ۵-۴ : واحد استخراج پاسخ در راهکار پیشنهادی.....۶۹

شکل ۶-۴ : واحد تحلیل رابطه در راهکار پیشنهادی.....۷۱

شکل ۱-۵ : خروجی تجزیه‌گر برای جمله.....۷۸

شکل ۲-۵ : فلوچارت پیاده‌سازی راهکار پیشنهادی.....۸۳

شکل ۱-۶ : معیار Pyramid-score در ۳۵ اجرا.....۹۱

فصل اول

مقدمه

زندگی بشر در گذار از ادوار تاریخ، اشکال مختلف به خود گرفت و در هر دوره نیازهای ویژه‌ای را پیش روی انسان نهاد، نیازهایی که در راستای رفع هر یک از آنها چالش‌های دیگری متولد می‌شد. بر این اساس تاریخ زندگی بشر را تاکنون می‌توان به سه دوره کشاورزی، صنعتی و اطلاعات تقسیم‌بندی کرد.

بسیاری از محققین، قرن بیستم و بیست و یکم را قرن اطلاعات می‌دانند. دوره‌ای که در آن بشر به دنبال ساز و کارهایی برای تولید، جمع‌آوری و پردازش اطلاعات است. نتیجه تلاش‌های انجام گرفته در این دوره، زمینه‌ساز مشکل انفجار اطلاعات بود. فعالیت‌های بشر در این دوره، وی را با چالش‌هایی برای مدیریت حجم عظیم اطلاعات مواجه کرد. مدیریت اطلاعات شامل ارائه راهکاری برای ذخیره‌سازی مؤثر، تعریف روش دسترسی و درنهایت ارائه ابزاری قوی برای جستجو در میان اطلاعات می‌باشد. مهم‌ترین ابزاری که در رویارویی با حجم بالای اطلاعات به یاری انسان می‌شتابد ابزار جستجو در میان اطلاعات است؛ اما تعریف هر روشی برای جستجو در میان اطلاعات وابسته به روش نمایش و ذخیره‌سازی اطلاعات است، که خود از نوع اطلاعات تأثیر می‌پذیرد.

اطلاعات را می‌توان به دو گروه ساختارمند^۱ و غیرساختارمند^۲ تقسیم کرد. اطلاعات ساختارمند اطلاعاتی هستند که می‌توان آن‌ها را در قالب چند عنصر (مبتنی بر یک ساختار داده-ای) تعریف کرد که رابطه مشخصی بین آن‌ها برقرار است؛ مانند جداول مربوط به اطلاعات کارکنان یک بانک که به ازای هر شخص سطری در جدول تعریف شده است و دارای مشخصاتی چون نام، نام خانوادگی، سن و... می‌باشد. در آغاز راهکارهای متعددی برای ذخیره‌سازی این نوع از اطلاعات تعریف شد. اولین گروه از این راهکارها را می‌توان ایده ذخیره‌سازی اطلاعات در فایل دانست. این راهکار قدرت چندانی در مواجهه با مشکلاتی چون بروز افزونگی، حجم زیاد کد برای دسترسی و دست‌کاری داده‌ها و احتمال بروز افزونگی نداشت. ایده مطرح شده بعدی ایده ذخیره‌سازی و مدیریت در سیستم مدیریت پایگاه‌های داده بود. در این روش داده‌ها به صورت عناصر کاملاً قانونمند ذخیره می‌شوند. این نحوه ذخیره‌سازی در کنار مزیت‌هایی چون سرعت دسترسی،

¹ structured

² unstructured

یکپارچه‌سازی نمایش داده‌ها، توانایی کنترل حجم بالای اطلاعات و حداقل افزونگی یک ایراد کلی دارند، آن هم پافشاری بر معرفی زبانی برای دسترسی و دست‌کاری داده که منجر به محدود شدن کاربران سیستم می‌شود؛ اما با وجود این ایراد راهکار بسیار قوی برای داده‌های ساختارمند به شمار می‌آید. داده‌های غیرساختارمند داده‌هایی هستند که در حالت کلی در قالب عناصر به هم مرتبط تعریف نمی‌شوند. دو روش برای مواجهه با این نوع داده‌ها وجود دارد. اولین روش تبدیل اطلاعات و داده‌های غیرساختارمند به اطلاعات ساختارمند است. برای نیل به این مقصود بیشتر از دانش افراد خبره استفاده می‌شود. به این معنا که با دانش افراد خبره، روابط پنهان بین اجزای داده‌های غیر-ساختارمند کشف‌شده، سپس این روابط در قالب اطلاعات ساختارمند ذخیره می‌شوند و در نهایت به صورت جداولی سازمان‌دهی شده مورد استفاده قرار می‌گیرند. اسناد متنی موجود در سیستم‌های اطلاعاتی نمونه‌ای از داده‌ها غیرساختارمند هستند با استفاده از راهکار مطرح‌شده می‌توان این اسناد را در قالب جداول اطلاعاتی موجود در پایگاه داده تعریف کرد؛ بر این اساس به صورت داده ساختارمند تعریف می‌شود. ایراد اصلی که به این راهکار می‌تواند گرفت از دست دادن بخش عظیمی از مفاهیم نهفته در داده‌های غیرساختارمند می‌باشد. برای جلوگیری از اتلاف مفاهیم در این نوع از داده‌ها راهکار دیگری مطرح شد. راهکار دوم ذخیره بدون تغییر داده‌ها در قالب اصلی خود می‌باشد، در این روش به جای نادیده گرفتن مفاهیم پنهان در این نوع از داده‌ها، سعی در کشف آن‌ها داریم. این خود عامل اصلی ظهور علم پردازش زبان طبیعی¹ بود. در این علم به دنبال تقویت این توانایی در کامپیوتر هستیم که بتواند متن را دریافت کرده آن را مطالعه کند و اطلاعات را خود از داخل متن استخراج کند.

در تمام راهکارهای ارائه‌شده برای نمایش، ذخیره‌سازی و جستجو در میان اطلاعات به دنبال پاسخ‌گویی به یک نیاز اساسی هستیم. نیازی که رویای بسیاری از محققین علوم کامپیوتر است. رویای تبدیل کامپیوتر به پیردانی که قادر به پاسخ‌گویی به هر سؤال باشد. در این راستا سیستم‌هایی ارائه شدند تا بتوانند با دریافت درخواستی از کاربر برای یک نیاز اطلاعاتی بهترین

¹ natural language processing

گزینه را ارائه دهند. در علم اطلاعات این سیستم‌ها را با نام سیستم‌های بازیابی اطلاعات می‌شناسیم. این نوع سیستم‌ها قادر هستند با دریافت نیاز اطلاعاتی از کاربر، بین اسناد موجود در سیستم به جستجو پرداخته، اسنادی که بیشترین تطبیق را با نیاز اطلاعاتی کاربر دارند ارائه کنند. نکته موجود در این سیستم این است که سیستم‌های بازیابی اطلاعات معمولاً با داده‌های غیر-ساختارمند سروکار دارند و تنها می‌توانند پیش‌بینی کنند که پاسخ به یک نیاز اطلاعاتی با چه احتمالی در هر یک از اسناد سیستم وجود دارد؛ اما توانایی پیدا کردن پاسخ کامل، جامع و درعین‌حال مختصر به نیاز اطلاعاتی کاربر را ندارند. پس هنوز رویای پیر دانا پابرجاست.

برای نزدیک شدن به تحقق این رویا، سیستم‌های دیگری تعریف شدند که این سیستم‌ها را با عنوان سیستم‌های پرسش و پاسخ یا به اختصار QA¹ می‌شناسیم. این سیستم‌ها می‌توانند یک نیاز اطلاعاتی را در قالب پرسشی از کاربر دریافت کرده و پاسخ را نیز در کمترین حجم ممکن به کاربر بازگردانند. این نوع سیستم‌ها روی داده‌های مختلفی چون متن، صوت، تصویر و چندرسانه‌ای قابل تعریف هستند. در ادامه این بخش ابتدا علت اهمیت طراحی چنین سیستم‌هایی را بیان کرده، سپس هدف اصلی ارائه این پژوهش را معرفی می‌کنیم و در انتها این فصل را با برشمردن مشکلات پیش رو در این پژوهش و معرفی ساختار این پایان‌نامه به پایان می‌بریم.

۱-۱ اهمیت موضوع

با همه‌گیر شدن تکنیک‌های دیجیتال برای انجام امور، زمان به چالش‌برانگیزترین پارامتر زندگی بشر تبدیل شده است، به‌گونه‌ای که استفاده از روش‌هایی که کمترین نیاز را به صرف وقت توسط انسان داشته باشد، اهمیت ویژه‌ای پیدا کرده است. پاسخ‌گویی به سؤالات هم از این قاعده مستثنا نیست. در این بین روش‌هایی که بتوانند با ارائه پاسخ مناسب به سؤالات، فرآیند استخراج پاسخ را تا حد قابل‌ملاحظه‌ای خودکار انجام دهد، می‌توانند نقش مهمی در صرفه‌جویی زمان داشته باشند.

¹ Question answering

۲-۱ فرضیات و هدف

شناخت مفاهیم و ارتباطات موجود بین آنها برای پاسخ به سؤال در هر زمینه‌ای لازم است. پاسخ‌گویی بدون درک مفاهیم و ویژگی‌ها و تنها با تکیه بر لغات مشابه هر سیستمی را دچار خطا می‌کند. باین‌وجود درک کلیه مفاهیم موجود در متون از توان سیستم‌های امروزی خارج است. علت اصلی این امر را می‌توان در ابهام موجود در زبان جستجو کرد. درحالی‌که ذهن انسان با شناختی که از مفاهیم دارد متون را تجزیه و تحلیل می‌کند و تا حدودی بر این ابهام غلبه می‌کند. انتظار می‌رود با آموزش روابط مفهومی موجود در حوزه‌های مختلف بتوان عملکرد سیستم‌های پرسش و پاسخ را از نقطه نظر منطق پاسخ‌گویی بهبود بخشید.

۳-۱ مراحل تحقیق و مشکلات

یکی از چالش‌برانگیزترین مراحل در انجام تحقیق در زمینه سیستم‌های پرسش و پاسخ انتخاب واحد بازیابی اطلاعات سیستم است. بر این اساس اولین مشکلی که در انجام این پژوهش مطرح شد، ایجاد ارتباط با موتورهای جستجو به عنوان واحد بازیابی اطلاعات بود. چراکه برای دسترسی به موتورهای جستجو نیازمند استفاده از آدرس شبکه‌های متفاوت است. دسترسی به آدرس شبکه‌های مختلف سرعت کارایی سیستم را به شدت کاهش می‌دهد و این فرآیند پاسخ به مجموعه سؤالات را زمان‌گیر می‌کند.

پس از طراحی ایده و پیاده‌سازی آن نوبت به تست روش می‌رسد. در این فاز نیاز به سؤالاتی برای تست سیستم مطرح می‌شود؛ اما مجموعه سؤال معتبری در حوزه فناوری اطلاعات و ارتباطات وجود نداشت؛ بنابراین فراهم آوردن این مجموعه یکی از چالش‌های پیش رو در این پژوهش به حساب می‌آید. برای رفع این مشکل از گروهی شامل یازده دانشجو و فارغ‌التحصیل رشته‌های مهندسی کامپیوتر و فناوری اطلاعات خواسته شد سؤالات خود را به همراه پاسخ مورد انتظار سیستم ارائه دهند. به این صورت مجموعه‌ای شامل ۱۳۲ سؤال تعریفی و سؤال پاسخ کوتاه گردآوری شده و در طول این تحقیق برای تست و ارزیابی سیستم مورد استفاده قرار گرفت.

۴-۱ ساختار پایان نامه

این پایان نامه در شش فصل تنظیم شده است. در فصل دوم ابتدا تعاریف موجود در دامنه تحقیق را بررسی کرده، سپس با مروری بر راهکارهای پیشین، این فصل را به پایان می‌رسانیم. در فصل سوم تعاریف انواع روابط زبانی و معنایی موجود بین لغات را معرفی می‌کنیم، این فصل مبانی موردنیاز در ارائه راهکار پیشنهادی را مطرح می‌کند. راهکار پیشنهادی در فصل چهارم به تفصیل مطرح شده، سپس در فصل پنجم به پیاده‌سازی راهکار پرداخته و چالش‌های آن را مورد بررسی قرار می‌دهیم. در نهایت مطلب را با مقایسه راهکار پیشنهادی با راهکارهای مشابه پیشین و بیان نقاط قوت و نقاط ضعف روش به پایان می‌بریم.

فصل دوم

مفاهیم اولیه و مروری بر کارهای گذشته

سیستم‌های پرسش و پاسخ، نسل جدید سیستم‌های بازیابی اطلاعات هستند، با این تفاوت که در پرسش و پاسخ به دنبال پاسخ‌گویی به یک سؤال در قالب یک کلمه، یک جمله و یا یک پاراگراف به جای بازگردانی کل یک سند هستیم. سیستم‌های بازیابی اطلاعات را می‌توان قلب تپنده سیستم‌های پرسش و پاسخ دانست. به این صورت که پرسش‌های ورودی بعد از پردازش اولیه وارد سیستم بازیابی اطلاعات شده و پاسخ‌ها در قالب چندین سند رتبه‌بندی شده از واحد بازیابی اطلاعات دریافت می‌شوند. پس از دریافت اسناد کاندید از واحد بازیابی اطلاعات به دنبال بهترین قسمت این اسناد برای پاسخ‌گویی به سؤال می‌باشیم. سیستمی که بتواند با دریافت سؤال از کاربر به جستجو در بین اسناد کاندید برای یافتن پاسخ بپردازد، نیازمند درک مفاهیم نهفته در پرسش‌های ورودی و اسناد کاندید هستیم. میزان و عمق این درک وابسته به روشی است که برای پاسخ‌گویی استفاده می‌شود. گاه این شناخت تنها به وجود و عدم وجود یک لغت در مجموعه اسناد بسنده می‌کند و گاهی تا حد درک کامل تمام روابط صرفی و نحوی موجود در اسناد پیش می‌رود.

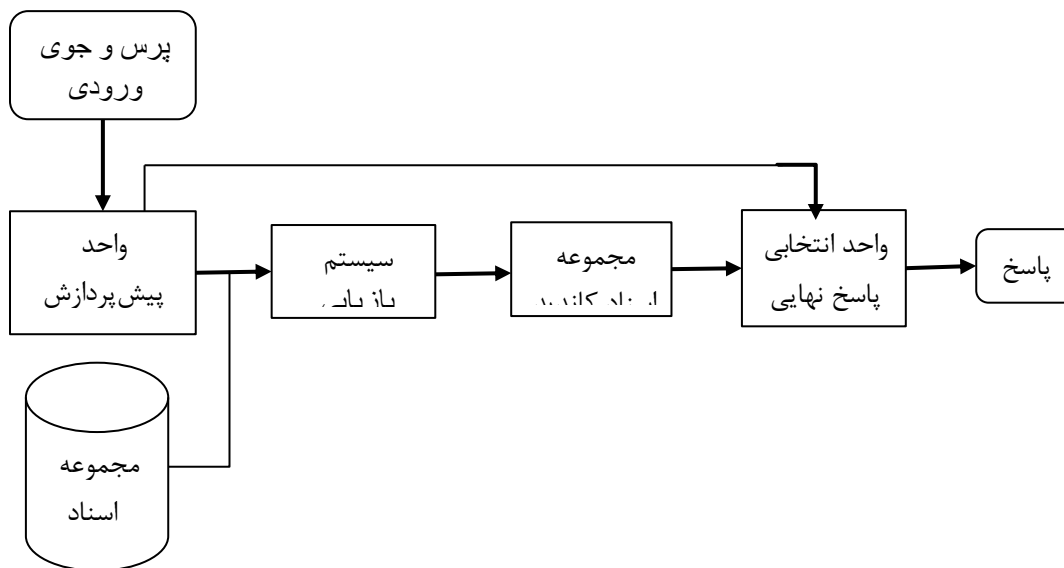
۲-۱ مفاهیم اولیه

سیستم‌های پرسش و پاسخ را از نظر سؤالاتی که قادر به پاسخ‌گویی هستند، می‌توان به دو دسته دامنه‌باز^۱ و دامنه‌بسته^۲ تقسیم‌بندی کرد. سیستم‌های پرسش و پاسخ دامنه‌بسته، سیستم‌هایی هستند که تنها به سؤالات از یک دامنه معنایی خاص پاسخ می‌دهند [۱]؛ مانند سیستمی که در سال ۲۰۰۹ برای پاسخ‌گویی به سؤالات حوزه زیست‌شناسی مطرح شد [۲]. در ایران هم نمونه‌هایی از سیستم‌های پرسش و پاسخ دامنه‌بسته در حوزه فروش کالا [۳] و زندگی‌نامه [۴] ارائه شده است. در مقابل سیستم‌های پرسش و پاسخ دامنه‌باز سیستم‌هایی هستند که در مورد سؤالات محدودیت موضوعی ندارند [۵] و پاسخ‌گوی سؤالات در تمام زمینه‌ها می‌باشند. نمونه‌ای از سیستم‌های پرسش و پاسخ دامنه‌باز در سال ۲۰۱۳ توسط فدر و همکارانش [۶] ارائه شد.

1 Open domain

2 Close domain

برای معرفی اصطلاحات رایج در سیستم‌های پرسش و پاسخ فرآیند کلی کار این نوع سیستم -
ها را مطابق شکل ۱-۲ نمایش می‌دهیم؛ اما معرفی معماری کلی را به بخش بعدی موکول می-
کنیم.



شکل ۱-۲: نمایش عملکرد کلی سیستم‌های پرسش و پاسخ

در سیستم‌های پرسش و پاسخ همه چیز با ورود یک نیاز اطلاعاتی و به درخواست کاربر آغاز می‌شود. کاربر نیاز اطلاعاتی خود را در قالب پرس‌وجو^۱ ارائه می‌دهد. پرس‌وجو عبارت است از بیان نمادین نیاز اطلاعاتی. در اکثر سیستم‌های پرسش و پاسخ یک پرس‌وجو می‌تواند در قالب زبان طبیعی به سیستم ارائه شود، حال آنکه در برخی موارد برای ارتباط با سیستم پرسش و پاسخ زبان خاصی تعریف می‌شود و کاربر باید پرس‌وجوی خود را در قالب این زبان تعریف شده به سیستم ارائه دهد. بعد از دریافت پرس‌وجوی ورودی ابتدا این پرس‌وجو تجزیه و تحلیل می‌شود. اعمالی که در فرآیند تجزیه و تحلیل یک پرس‌وجو انجام می‌شود، کاملاً وابسته به روشی است که برای بازیابی استفاده می‌شود. به عنوان مثال در برخی از روش‌های بازیابی اطلاعات هیچ حساسیتی نسبت به کلمات پرتکرار مانند "است"، "بود"، "از"، "اگر" و... وجود ندارد. در سیستم‌های پرسش و پاسخ

¹ query