قران‌جوی: اولین سامانه پرسش و پاسخ خودکار قرانی

مژگان فرهودی[۱]، احسان درودی[۲]، احسان شرکت[۳]، علی رهنما[۴]

پژوهشگاه ارتباطات و فناوری اطلاعات (مرکز تحقیقات مخابرات ایران)
{[1]farhoodi, [2]darrudi, [3] ehsansherkat, [4]arahnama}@itrc.ac.ir

چکیده

در این مقاله به معرفی و روش ساخت "قران‌جوی" به عنوان اولین سامانه پرسش و پاسخ قرانی به زبان فارسی می‌پردازیم. مبحث پرسش و پاسخ در قران در حوزه‌های ترجمه، مفاهیم و تفسیر قران از اهمیت بالایی برخوردار است. سامانه‌های متداول فعلی بازیابی اطلاعات، بر اساس یک یا چند کلمه کلیدی که کاربر وارد می‌کند، تعدادی سند را برمی‌گردانند. در این سامانه هدف آن است که ابتدا سایت‌ها و بانک‌های اطلاعاتی معتبر قرانی شناسایی شده و اطلاعات آن در یک مخزن اطلاعاتی ذخیره گردد. سپس سامانه پس از دریافت پرسش کاربر و پردازش آن (تعیین نوع پرسش، تأکید پرسش و استخراج کلمات کلیدی)، اسناد و اطلاعات موجود در مخزن داده را جستجو می‌کند و پس از یافتن پاسخ‌های مناسب (که ممکن است یک پاراگراف، یک جمله و یا یک عبارت باشد)، چند پاسخ برتر که رتبه بالاتری را کسب کرده‌اند به کاربر باز می‌گرداند. البته در کنار این مخزن اطلاعاتی، از یک هستان شناسی قرانی (بنام قران نگار که تنها هستان شناسی معتبر قرانی به زبان فارسی است و در همین طرح توسعه داده شده است) استفاده می‌گردد که این منجر به یافتن پاسخ های کوتاه و دقیق از منابع دانش سامانه می شود. نتایج حاکی از آن است که سامانه با دقت خوبی می‌تواند پاسخ‌های مورد نیاز کاربران را بازگرداند.

کلمات کلیدی

سامانه پرسش و پاسخ قرانی، قران‌جوی، گراف مفاهیم قرانی، پردازش پرسش، استخراج پاسخ

۱- مقدمه

در دهه‌های اخیر حجم اطلاعات تولیدشده توسط بشر به صورت تصاعدی بالا رفته است، به نحوی که پیدا کردن دانش مورد نیاز از انبوه داده‌های موجود فعالیتی چالش برانگیز شده است. برای رفع این مشکل، سامانه‌های بازیابی اطلاعات [Baeza 1999] طراحی شدند که هم اکنون نیز انواع تحت وب آنها (مانند گوگل) محبوبیت فراوانی دارند. سامانه‌های پرسش و پاسخ نوع عالی‌تری از سامانه‌های بازیابی اطلاعات (مانند موتور جستجوی گوگل) محسوب می‌شوند که یک پرسش را به زبان طبیعی از کاربر دریافت نموده و سپس جواب را از منبع دانش خود استخراج می‌نماید و به کاربر برمی‌گردانند [Dwivedi 2012].

در مقایسه با سامانه‌های بازیابی اطلاعات کلاسیک که در آنها واحد ارائه اطلاعات سند است سامانه‌های پرسش و پاسخ تلاش می‌نمایند «جواب دقیق» را مستقیما محاسبه و ارائه نمایند که باعث تسهیل و افزایش سرعت دسترسی به دانش می‌شود. این امکان برای برخی کاربردهای آینده‌نگر که نیاز به تصمیم‌گیری‌های بلادرنگ دارند حیاتی است. همچنین چون روش ارتباط کاربر با این سامانه به گفتمان انسان نزدیک‌تر است امکان تعامل‌پذیری[۱] آنها بالاتر می‌باشد. هرچند ارائه این

قابلیت‌ها، چالش‌های بسیار بزرگی مانند فهم نسبی جملات زبان طبیعی را به دنبال دارد.

در این مقاله به ارائه یک سامانه پرسش و پاسخ فارسی مخصوص حوزه قرآنی با نام "قران‌جوی" می‌پردازیم. این سامانه، اولین سامانه پرسش و پاسخ کاملا خودکار برای این حوزه می‌باشد و تا جایی که اطلاع داریم اولین سامانه پرسش و پاسخ مقیاس‌پذیر[۲] برای زبان فارسی می‌باشد. این سامانه از طریق وبگاه اینترنتی[۳] جهت استفاده عموم در دسترس می‌باشد.

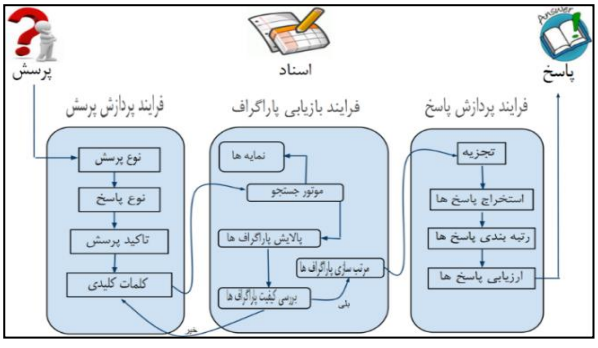
در ادامه، این نوشتار به هفت بخش اصلی تقسیم شده است. در بخش ۲ به شرح کارهای مشابه پرداخته شده است. در بخش ۳ معماری سامانه بیان شده و در بخش‌های ۴ و ۵ به ترتیب به شرح گراف مفاهیم قرانی و فرایندهای سامانه می‌پردازیم. در نهایت در بخش۶ نتایج آزمایشات بخش‌های مختلف سامانه ذکر شده است. در بخش ۷ نیز جمع‌بندی نهایی آورده شده است.

۲- کارهای مرتبط

سامانه‌های پرسش و پاسخ مختلفی در سال‌های اخیر چه به صورت آزمایشگاهی و چه به صورت تجاری ایجاد شده است. از آنجایی که سامانه‌های پرسش و پاسخ می‌توانند رقیبی جدی برای موتورهای جستجو محسوب شوند، تمامی شرکت‌های بزرگ نرم‌افزاری از جمله ای‌بی‌ام[۴] و یاهو[۵] در آزمایشگاه‌های خود بر روی این سامانه‌ها تحقیقات گسترده‌ای انجام می‌دهند. بررسی سامانه‌های مختلف پرسش و پاسخ چه در دامنه باز و چه در دامنه محدود، می‌تواند در تعیین معماری و ابزارهای مختلف مورد نیاز برای سامانه پرسش و پاسخ قرآنی راه‌گشا باشد.

اولین تلاش‌ها برای ایجاد سامانه‌های پرسش و پاسخ مربوط به سال ۱۹۶۴ و سامانه الیزا[۶] می‌باشد (Weizenbaum 1966). این سامانه نمونه اولین سامانه‌های پردازش زبان طبیعی می‌باشد که برای تعامل و پاسخ دادن به صحبت‌های کاربران به صورت متنی در دانشگاه ام‌ای‌تی[۷] ایجاد شد. براساس الیزا سامانه داکتر[۸] ایجاد شد که یک برنامه کامپیوتری برای تعامل با کاربران برای روان‌درمانی می‌باشد. سامانه دیگری به نام لونار[۹] نیز بر اساس الیزا بنا شده است. این سامانه برای پاسخگویی به سؤالات در دامنه علوم فضایی می‌باشد و قادر است به ۹۰ درصد از سؤالات پاسخ صحیح بدهد. در دهه ۶۰ عمدتاً سامانه‌های پرسش و پاسخ در تست تورینگ[۱۰] مورد بررسی قرار می‌گرفتند.

نمونه‌های بسیار دیگری نیز مورد بررسی قرار گرفتند. به عنوان مثال AnswerBus یک سامانه پرسش و پاسخ دامنه باز[۱۱] می‌باشد که برای پرسش‌های کاربران پاسخ‌هایی در حد جمله پیدا می‌کند. این سامانه قادر است پرسش کاربران را در زبان‌های انگلیسی، آلمانی، فرانسوی، اسپانیایی، ایتالیایی و پرتغالی دریافت کند و در چند ثانیه برای آن‌ها پاسخ‌هایی به زبان انگلیسی فراهم کند (Zheng 2002). AQUA یک سامانه پرسش و پاسخ است که در دانشگاه آزاد کشور انگلستان توسعه پیدا کرده است (Vargas-Vera and Motta 2004). یکی از مشخصه‌های اصلی این سامانه استفاده از دانش (ذخیره شده به صورت هستان‌شناسی) می‌باشد. این دانش در مراحل مختلف پرسش و پاسخ مانند اصلاح پرسش بکار برده می‌شود. سامانه TelQAS می‌تواند به پرسش‌های مطرح شده در حوزه تخصصی مخابرات پاسخ دهد (حجازی و همکاران ۱۳۸۳) . پایگاه داده سامانه حاوی مفاهیم و واقعیات اطلاعات مربوط به حوزه تخصصی مورد نظر است که در قالب گراف هستان‌شناسی ذخیره شده است. OpenEphyra اولین چارچوب[۱۲] باز برای سامانه‌های پرسش و پاسخ می‌باشد (Schlaefer et al. 2006). چارچوب ارائه شده می‌تواند علاوه بر زبان انگلیسی بر روی سایر زبان‌ها نیز با جایگذاری اجزای زبانی اعمال شود. این سامانه از وب برای پایگاه دانش خود استفاده می‌نماید، ولی قادر است از مجموعه‌های داده‌ای کوچک‌تر نیز استفاده کند. بنابراین این سامانه بیشتر برای دامنه باز طراحی شده است هر چند که می‌توان از آن برای دامنه محدود نیز استفاده نمود. سامانه پرسش و پاسخ Mulder در سال ۲۰۰۱ در دانشگاه واشنگتن ساخته شده است (Kwok et al. 2001). هدف اصلی از طراحی این سامانه گسترش دامنه سامانه‌های پرسش و پاسخ به دامنه وب می‌باشد به طوری که به ادعای سازندگان این سامانه، این اولین سامانه پرسش و پاسخ با دامنه باز می‌باشد. سامانه پرسش و پاسخ تک زبانه اسپانیایی(Vicedo et al. 2004) در بخش پردازش‌های زبان طبیعی و سامانه‌های اطلاعاتی دانشگاه آلینتکا[۱۳] ساخته شده است. این سامانه بر پایه ترکیب ابزارهای کم‌عمق پردازش زبان‌های طبیعی و تکنیک‌های حذف آماری ساخته شده است. سامانه تنها قادر است روی اسناد اسپانیایی کار کند و می‌تواند برای پرسش‌ها پاسخ‌های دقیق بیابد. QArab سامانه پرسش و پاسخی است که پرسش‌ها را به زبان عربی گرفته و تلاش می‌کند تا برای آن‌ها پاسخ کوتاهی بیابد. منبع اطلاعات اولیه این سامانه، مجموعه‌ای از متون استخراج شده از روزنامه "الرایا" می‌باشد. در این سامانه از موتورهای جستجوی سنتی برای بازیابی اطلاعات استفاده می‌شود، اما بر روی پرسش و پاسخ‌های یافت شده پردازش‌های زبانی صورت می‌گیرد. سامانه پرسش و پاسخ هواشناسی یک مدل پرسش و پاسخ عمومی است که بر روی یک دامنه محدود (هواشناسی) و بر پایه الگوهای تعریف شده به پیش‌بینی وضع هوا می‌پردازد. مدل طراحی شده برای



شکل۱: معماری کلی یک سامانه پرسش و پاسخ

تحقق هدف استخراج پاسخ در حداقل زمان از روابط معنایی و نحوی، پرسش‌های قبلی و الگوهای پویا استفاده می‌کند. اگر هیچ الگوی مناسبی برای پاسخ یافت نشود کاربر می‌تواند الگوی مناسبی با توجه به گرامر زبان انگلیسی بسازد. سامانه AquaLog یک سامانه قابل حمل[۱۴] پرسش و پاسخ می‌باشد، به طوری که پرسش‌های کاربران به صورت زبان طبیعی و هستان‌شناسی مربوطه را از ورودی دریافت، سپس جواب‌هایی را بر اساس پایگاه دانش خود به سؤالات پرسیده شده به کاربران نشان می‌دهد (Lopez et al. 2005a). PowerAqua یک سامانه پرسش و پاسخ مبتنی بر چند هستان شناسی[۱۵] است که پرس-وجوهای زبان طبیعی را به عنوان ورودی دریافت می‌کند و این توانایی را دارد تا از منابع توزیع شده مرتبطِ موجود در وب معنایی[۱۶]، پاسخ مناسب را استخراج کند. لازم به ذکر است که PowerAqua فقط به یک هستان شناسی محدود نمی‌شود. PowerAqua سامانه جدیدی است که برای پشتیبانی از پرسش و پاسخ در حوزه وب معنایی، قابلیت‌های بیشتری نسبت به AquqLog دارد (lopez et al. 2005). سامانه DeepQA توسط گروه تحقیقاتی IBM ساخته شده است با این هدف که بتواند با یک انسان به صورت بلادرنگ[۱۷] در یک مسابقه‌ی تلوزیونی، Jeopardy، به رقابت بپردازد. این مسابقه دامنه‌ی وسیعی از موضوعات را پوشش می‌دهد. این نیازمندی باعث طراحی معماری DeepQA و پیاده‌سازی واتسون[۱۸] شده است.

۳- معماری سامانه

سامانه‌های پرسش و پاسخ در حالت کلی، گونه‌ای از سامانه‌های بازیابی اطلاعات بشمار می‌روند که با در اختیار داشتن مجموعه‌ای از اسناد، می‌کوشند تا برای پرسش‌های مطرح که اغلب در قالب زبان طبیعی هستند، پاسخ‌های مناسب را استخراج نمایند. بر این اساس، این قبیل سامانه‌ها نسبت به سامانه‌های معمول بازیابی اطلاعات، به تکنیک‌های پیچیده‌تر پردازش زبان طبیعی[۱۹] سروکار دارند و در محافل علمی به عنوان نسل آینده موتورهای جستجوی اطلاعات مطرح هستند.

۳-۱- معماری کلی سامانه‌های پرسش و پاسخ

ابتدا باید خاطر نشان کرد سامانه‌های پرسش و پاسخ حوزه محدود که بر اساس منابع دانش و نیازمندی‌های خاص خود طراحی و پیاده‌سازی می‌شوند بعضا از روش‌های تک منظوره استفاده می‌نمایند، لذا معماری مشترکی برای این سامانه‌ها نمی‌توان متصور شد (Unger 2013). لیکن سامانه‌های حوزه باز چون تعریف و منابع دانش یکسانی دارند (متون خام) معمولا از یک معماری مشترک تبعیت می‌نمایند که در شکل ۱ نشان داده شده است.

شکل۱ یک معماری کلی از سامانه پرسش و پاسخ می‌باشد که از مقاله لمپرت[۲۰] (Lampert 2004) با کمی اصلاح اقتباس شده است.

در ادامه به تشریح بخش‌های مختلف این معماری پرداخته می‌شود.

- فرایند پردازش پرسش[۲۱]:

فرایند پردازش پرسش اولین مرحله فرایند پرسش و پاسخ می-باشد. این فرایند پرسش کاربر را به عنوان ورودی دریافت می‌کند و پردازش‌های لازم را روی آن انجام می‌دهد تا برای فرایندهای بعدی آماده شوند؛ این فرایندشامل بخش‌های زیر است (فرهودی و سموری ۱۳۸۹):

الف) طبقه‌بندی نوع پرسش/پاسخ[۲۲]: برای یافتن پاسخ مناسب در ابتدا باید بدانیم که دنبال چه چیزهایی در میان اسناد بگردیم. نوع پرسش به ما کمک می‌کند که اطلاعات مرتبط را محدودتر کنیم و موجب تسهیل فرآیندهای بازیابی پاراگراف و پردازش پرسش می‌شود. نوع پاسخ بر مبنای نوع پرسش انتخاب می‌شود، بنابراین طبقه‌بندی پرسش بخش مهمی از پردازش پرسش می‌باشد و حتی اگر به عنوان یک بخش اساسی مدنظر نباشد می‌تواند اطلاعات مهمی را برای پاسخ به پرسش کاربر ارائه کند.

ب) تعیین تأکید پرسش[۲۳]: گاهی دانستن نوع پرسش به تنهایی برای پیدا کردن پاسخ کفایت نمی‌کند خصوصاً اینکه بعضی از پرسش‌ها می‌توانند بسیار ابهام آمیز باشند؛ به منظور کاهش ابهام به یک بخش اضافی برای پیدا کردن تأکید پرسش احتیاج است. منظور از تأکید پرسش، کلمه یا مجموعه کلماتی در پرسش می‌باشد که اشاره به اطلاعات درخواستی توسط پرسش دارند. برای مثال در پرسش "چه کسی به مدرسه رفت؟" تأکید پرسش روی "مدرسه" می‌باشد. اگر نوع پرسش و تأکید آن مشخص شود، سامانه راحت‌تر می‌تواند نوع پاسخ را تشخیص بدهد.

ج) استخراج کلمات کلیدی[۲۴]: در این قسمت کلمات کلیدی متناظر با پرسش کاربر بر حسب اولویتی که دارند به فرایند بازیابی اسناد فرستاده می‌شوند و کلمات بی‌ارزش حذف می‌شوند.

- فرایند بازیابی پاراگراف[۲۵]:

این فرایند بر مبنای یک یا چند سامانه مجزای بازیابی اطلاعات به جمع‌آوری اسناد مرتبط از میان اسناد موجود در وب می‌پردازد، سپس نتایج حاصل از سامانه بازیابی به منظور حذف قسمت‌هایی که شامل کلمات کلیدی پرسش نمی‌باشند مورد پالایش قرار می‌گیرند.

بعد از ارزیابی کیفیت پاراگراف‌ها، آنها به ترتیب احتمال قرار گرفتن پاسخ مرتب می‌شوند. اگر پاراگراف‌ها خیلی کم یا زیاد باشند از کاربر خواسته می‌شود تا پرسش خود ر ا با کاهش یا افزایش کلمات کلیدی تغییر دهد و دوباره ارسال کند. این کار تضمین می‌کند تعداد قابل قبولی از اسناد به قسمت بعدی فرستاده شود. انگیزه کاهش اسناد به پاراگراف قبل از فرستادن به فرایند پردازش پرسش منجر به تسهیل و تسریع فعالیت‌های فرایند بعدی می‌شود. فرایند مذکور شامل قسمت‌های زیر می‌باشد(Vicedo et al. 2004):

الف) بازیابی اسناد: در این قسمت سامانه پرسش و پاسخ از یک موتور جستجو برای بازیابی بهترین اسناد مرتبط با پرسش کاربر استفاده می‌کند.

ب) پالایش پاراگراف[۲۶]: همانطور که بیان شد تعداد اسنادی که توسط موتور جستجو برگردانده می‌شود بسیار زیاد است. بخش پالایش پاراگراف به منظور کاهش تعداد اسناد و کاهش متن هر سند به کار برده می‌شود.

ج) بررسی کیفت پاراگراف[۲۷]: این بخش وظیفه ارزیابی کیفیت پاراگراف‌های کاندید را دارد. اگر کیفیت پاراگراف‌ها مناسب نباشد سامانه مجدداً به بخش استخراج کلمات کلیدی برمی‌گردد تا کلمات کلیدی جدیدی را انتخاب نماید.

د) مرتب‌سازی پاراگراف‌ها[۲۸]: هدف این است که پاراگراف‌ها را با توجه به شانس دارا بودن پاسخ رتبه‌بندی کنند.

- فرایند پردازش پاسخ[۲۹]:

آخرین فرایند معماری سامانه پرسش و پاس، پردازش پاسخ است که وظیفه تشخیص و استخراج پاسخ‌ها را از میان پاراگراف‌های منتخب ایفا می‌کند و می‌تواند شامل بخش‌های زیر باشد:

الف) استخراج پاسخ‌ها[۳۰]:پاراگراف‌های منتخب برای استخراج پاسخ وارد این بخش می‌شوند. این بخش می‌تواند از نوع پاسخ که در مرحله پردازش پرسش به دست آمده و همچنین از کلمات کلیدی درون پرسش برای استخراج پاسخ‌ها کمک بگیرد(Kangavari et al. 2008).

ب) رتبه‌بندی پاسخ‌ها[۳۱]: این بخش پاسخ‌ها را دریافت کرده و آنها را به ترتیب اولویت مرتب می‌کند. اولویت هر پاسخ می‌تواند بر مبنای تکرار نوع پاسخ و همچنین چگالی کلمات کلیدی موجود در آن باشد(Kangavari et al. 2008) .

ج) صحت پاسخ[۳۲]ها: اعتماد به درستی پاسخ را می‌توان از روش‌های مختلفی بالا برد. یک روش استفاده از منابع لغوی مانند وردنت برای بررسی درستی پاسخ کاندید می‌باشد و یک راه دیگر استفاده از پایگاه دانش برای بررسی درستی پاسخ است. نظارت کاربران نیز می‌تواند برای این منظور به کار برده شود.

۳-۲- معماری پیشنهادی سامانه پرسش و پاسخ قرانی (قران‌جوی)

برای طراحی معماری سامانه پرسش و پاسخ به نحوی که بتواند به سوالات کاربر با کیفیت مناسب پاسخگو باشد، نکاتی در نظر گرفته شده است که این طراحی را مناسب با یک سامانه کامل پرسش و پاسخ می‌کند. اهم این موارد در ذیل آمده است:

۱. در یک سامانه پرسش و پاسخ می‌توان از منبع‌های دانش متفاوتی بهره برد (هستان‌شناسی، پرسش و پاسخ‌های متداول (FAQ)، پایگاه‌داده و غیره). به عنوان مثال، اگر بتوان یک هستان‌شناسی یا گراف مفاهیم قرآنی تهیه کرد، می‌توان به سادگی آنرا در قسمت منابع دانش وارد کرد و از مزایای آن در پروژه بهره برد.

۲. اجزای سامانه پرسش و پاسخ قرآنی به نحوی در این طراحی در نظر گرفته شده‌اند که مستقل از یکدیگر عمل می‌کنند. به عنوان مثال در صورتی که الگوریتم پردازش یا تحلیل پرسش تغییر کند، تأثیر بر مؤلفه بازیابی و پردازش ند نخواهد داشت. همچنین در صورتی که هر یک از منابع دانش بهبود یابند، نیازی به تغییر سایر اجزای سامانه نخواهد بود.

۳. در صورت نیاز، برای توسعه سامانه می‌توان به سادگی مولفه‌های دیگر استخراج پاسخ را به سامانه اضافه نمود.

شکل ۲، چارچوب مفهومی پیشنهادی سامانه پرسش و پاسخ قرآنی را نشان می‌دهد. همانطور که در شکل مشخص است، این چارچوب از سه بخش اصلی زیر تشکیل شده است:

- فرآیند برخط
- فرآیند برون‌خط
- رابط منابع دانش

فرآیند برخطاین بخش وظیفه اصلی انجام پرسش و پاسخ و تعامل با کاربر را دارد. الگوریتم‌های پرسش و پاسخ در این بخش اجرا می‌شوند و شامل مؤلفه¬¬های زیر است:

الف) تحلیل پرسش

وظیفه این مؤلفه آن است که پرسش کاربر که به زبان طبیعی وارد سامانه شده است را پردازش کرده و سپس اطلاعات مورد نیاز مؤلفه‌های بعدی سامانه را از آن استخراج می‌کند.

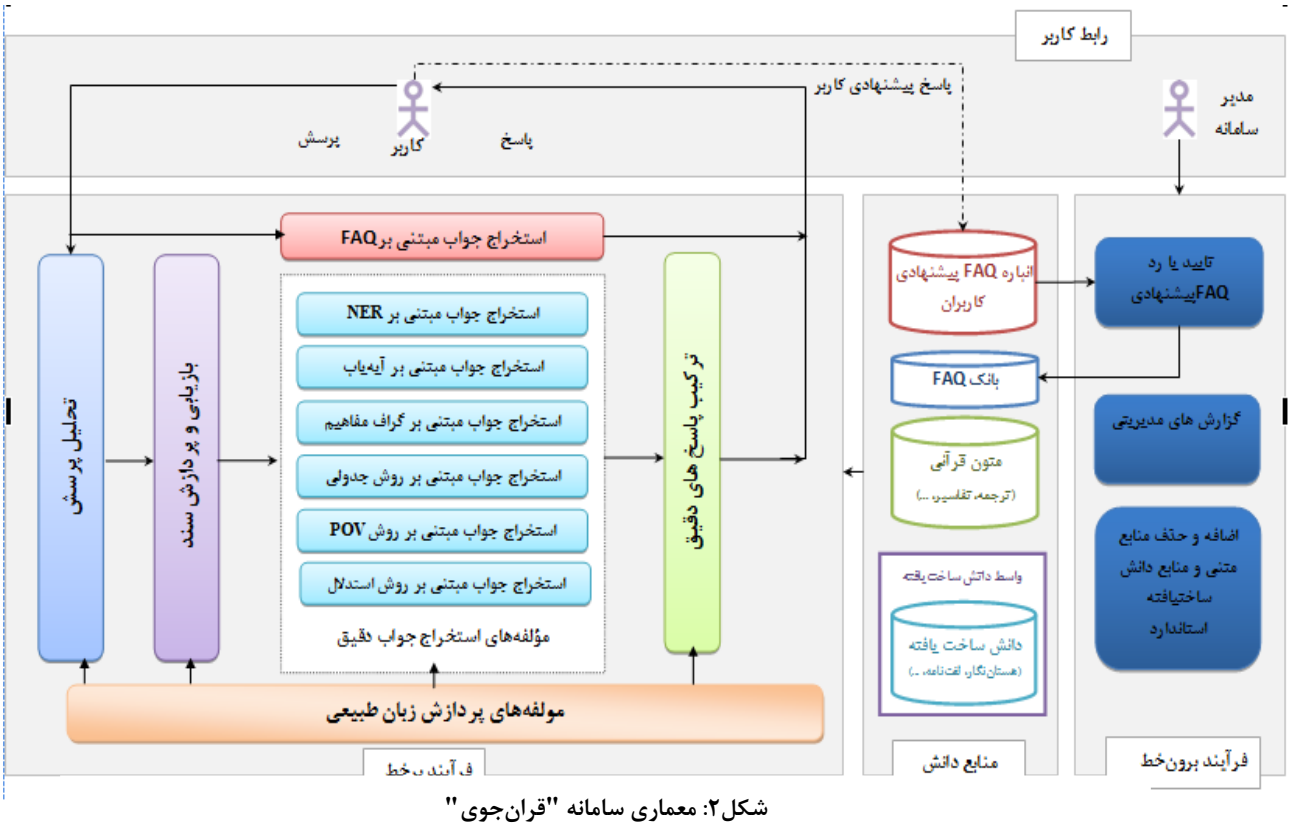
ورودی:

- پرسش به زبان طبیعی

خروجی:

خروجی «پرسش تحلیل شده» نام دارد که شامل بازنمایی‌های پرسش در قالب مورد نیاز مؤلفه‌های بعدی است. مؤلفه تحلیل پرسش حداکثر تلاش خود را خواهد کرد تا این بازنمایی‌ها را استخراج نماید، چراکه موفقیت مؤلفه‌های بعدی دقیقاً به فهم و بازنمایی صحیح پرسش وابسته است. مسلماً اگر یک نوع بازنمایی قابل استخراج نباشد در مراحل بعدی مؤلفه متناظر آن امکان اجرا و تولید پاسخ را نخواهد داشت. اگر برای هر نوع بازنمایی چند خروجی تولید شود مؤلفه باید بتواند به هر بازنمایی یک درصد اطمینان نیز اختصاص دهد تا راهنمای مؤلفه‌های بعدی باشد(Sherkat 2014) .

همچنین در «پرسش تحلیل شده» خروجی مؤلفه‌های پردازش زبان طبیعی نیز گنجانده می‌شود تا کار برای مؤلفه‌های بعدی ساده‌تر و سریعتر انجام شود.



شکل۲: معماری سامانه "قران‌جوی"

ب) بازیابی و پردازش سند

این مؤلفه، فرایند بازیابی اطلاعات استاندارد را انجام می‌دهد. به عبارت دیگر، بر اساس کلمات کلیدی پرسش، اسناد مرتبط را از منابع متنی سامانه استخراج می‌نماید (واحد بازیابی می‌تواند پاراگراف یا جمله باشد). سپس پاراگراف‌های بازیابی شده را با مؤلفه‌های پردازش زبان طبیعی تحلیل می‌نماید.

ورودی:

- پرسش تحلیل شده
- منابع متنی پردازش شده (نمایه‌سازی شده)

خروجی:

خروجی «پاراگراف تحلیل شده» نام دارد و شامل پاراگراف‌های بازیابی شده و بازنمایی آنها در قالب پردازش شده توسط مولفه‌های پردازش زبان طبیعی خواهد بود.

اصل پاراگراف به زبان طبیعی مستقیماً به مؤلفه فرموله نمودن پاسخ می‌روند تا به شکل مناسب به کاربر نشان داده شوند. از طرف دیگر پاراگراف‌های تحلیل شده به عنوان ورودی مؤلفه‌های استخراج جواب دقیق مورد استفاده قرار می‌گیرند.

ج) مؤلفه استخراج جواب دقیق مبتنی بر FAQ

این مؤلفه تلاش می‌نماید FAQ منطبق با پرسش وارد شده را از بانک FAQ سامانه استخراج نماید. این پاسخ‌ها به مؤلفه «ترکیب پاسخ های دقیق» می‌روند تا با سایر پاسخ‌های دقیق تولید شده توسط سایر مؤلفه‌ها ترکیب شوند.

ورودی:

- پرسش تحلیل شده

خروجی:

- جواب‌های دقیق به همراه میزان درصد اطمینان به آنها

د) مؤلفه‌های استخراج جواب دقیق

این مؤلفه‌ها بر اساس پرسش و پاراگراف تحلیل شده با الگوریتم‌های خاص خود اقدام به ردیابی و استخراج جواب دقیق از پاراگراف می‌نمایند.

ورودی:

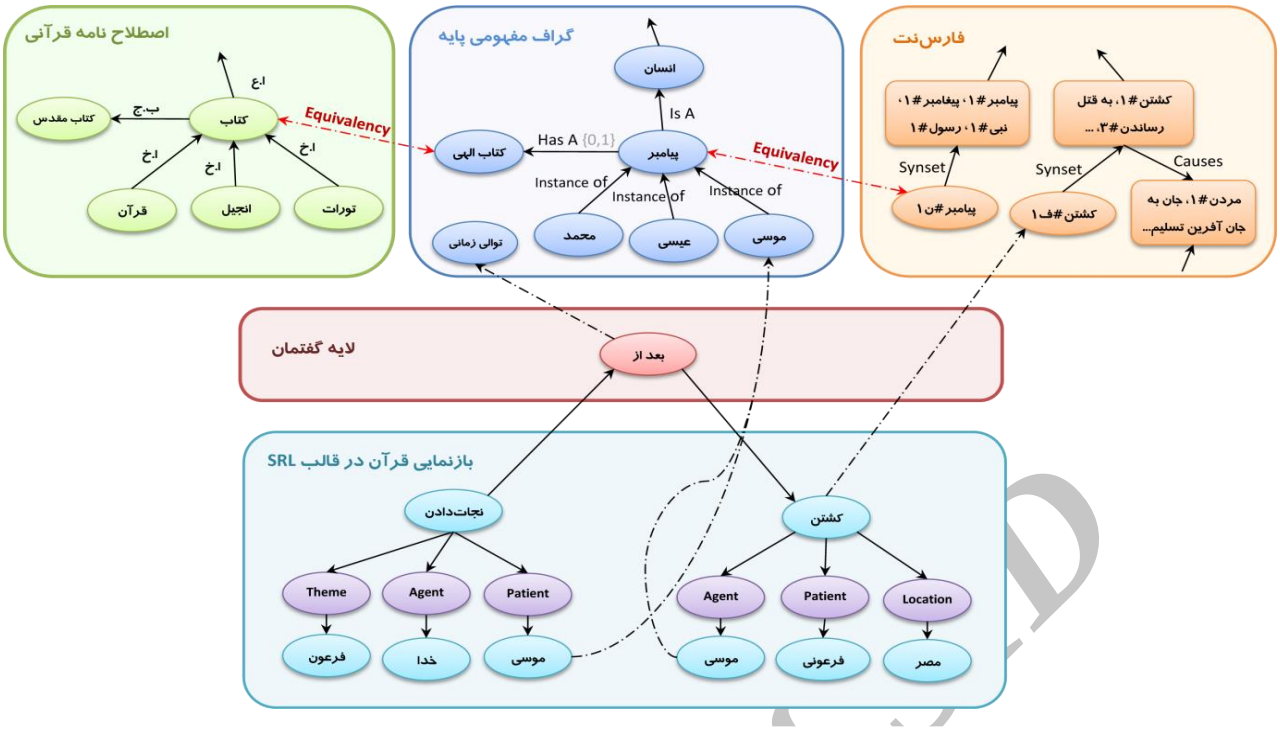
- پاراگراف‌های تحلیل شده
- منابع متنی پردازش شده (نمایه‌سازی شده)
- گراف مفاهیم قرانی

خروجی:

- جواب‌های دقیق به همراه میزان درصد اطمینان به آنها

ه) مؤلفه ترکیب پرسش‌های دقیق

مؤلفه‌های تولید پاسخ دقیق ممکن است تعداد زیادی پاسخ تولید نمایند ولی مسلماً نشان دادن همه آنها به کاربر صحیح نمی‌باشد. مؤلفه ترکیب پاسخ‌های دقیق باید بتواند بر اساس درصد احتمالی که هر یک از مؤلفه‌های قبلی به پاسخ‌های خود داده‌اند و همچنین میزان دقت



شکل ۳: گراف مفاهیم قرانی

خود مؤلفه‌ها، پاسخ‌ها را ترکیب نموده و چند پاسخ برتر را انتخاب نماید. به عنوان مثال پاسخی که هم توسط مؤلفه استدلال و هم توسط مؤلفه مبتنی بر الگو استخراج شده باشد مسلماً امتیاز بالاتری خواهد گرفت.

ورودی

- پاسخ های دقیق ارسالی از مؤلفه‌های استخراج پاسخ دقیق

خروجی

- چند پاسخ دقیق برتر

۳-۱-۱- فرآیند برون‌خط

این بخش شامل وظایف مدیریتی و نگهداری سامانه است که توسط مدیر وبگاه انجام می‌شوند. این وظایف مستقل از فرآیند برخط انجام می‌شوند و عبارتند از:

۱. بررسی انباره FAQ های وارد شده توسط کاربران سامانه و اضافه نمودن موارد مناسب به بانک FAQ دائم
۲. اضافه و حذف منابع دانش سامانه شامل منابع متنی و ساخت یافته
۳. تولید گزارش‌های مدیریتی از نحوه عملکرد سامانه

۳-۱-۲- رابط منابع دانش

منابع دانش بین فرآیندهای برخط و برون‌خط مشترک هستند و شامل قطعات زیر است:

۱. انباره FQA های وارد شده توسط کاربران سامانه
۲. بانک FAQ دائم
۳. متون قرآنی شامل ترجمه قرآن، تفاسیر، اصل قرآن به زبان عربی و غیره.
۴. منابع دانش ساخت‌یافته شامل هستان‌شناسی یا گراف مفاهیم حوزه قرآنی، فارس‌نت، وردنت، بانک‌های داده شامل لیست‌ها و جداول تعداد رخداد آیات و کلمات قرآن و غیره.

۴- گراف مفاهیم قرانی

همان طور که در شکل ۲ مشخص است سامانه قران‌جوی از روش‌ها و مؤلفه‌های گوناگون جهت استخراج جواب استفاده می‌نماید. به عنوان مثال منبع دانش مورد نیاز مؤلفه‌هایی مانند استخراج پاسخ مبتنی بر NER، متون معمولی هستند ولی برای سایر مؤلفه‌ها که در سطوح بالاتر معنایی کار می‌کنند لازم است از منابع دانش ساخت‌یافته استفاده شود. یکی از این مؤلفه‌ها موتور استدلال[۳۳] می‌باشد که سعی می‌نماید پاسخ پرسش تحلیل شده را از درون منابع دانش ساخت‌یافته سامانه استنتاج نماید. علاوه بر موتور استدلال، مؤلفه‌های تحلیل پرسش، ترکیب پاسخ و برخی مؤلفه‌های استخراج پاسخ نیز از دانش ساخت‌یافته استفاده می‌برند. لذا ساخت یک گراف مفاهیم قرانی از نیازمندی‌های این سامانه می‌باشد.

برای ساخت گراف پایه مفاهیم قرانی، نمونه‌های مشابه مورد بررسی قرار گرفت. در این خصوص فعالیت‌های پراکنده‌ای انجام شده بود ولی بدلایل مختلف (حجم بسیار کم و عدم جامعیت آنها، روابط و

مفاهیم بسیار محدود) پاسخگوی نیاز سامانه پرسش و پاسخ نبود. لذا ما اقدام به ایجاد اولین گراف مفاهیم حوزه قرانی نمودیم. برای شروع، هستان‌شناسی Leeds[۳۴] به عنوان پایه قرار گرفت. گراف مفهومی قرانی لیدز به زبان انگلیسی و عربی بوده و نیاز داشت تا به فارسی تبدیل شود. گراف مفاهیم لیدز بسیار محدود بود و تنها ۶۵ مفهوم قرانی را شامل می‌شد، لذا ما برای توسعه گراف از منابع معتبر قرانی از قبیل فرهنگ قران (شامل ۲۴ جلد)، تفاسیر، متن ترجمه قران و غیره استفاده نمودیم. لازم به ذکر است تمامی مفاهیم و روابط وارد شده در گراف به تایید کارشناسان قرانی رسیده است. جدول زیر آمار گراف پایه تولید شده را نشان می‌دهد:

جدول۱: آمار گراف مفاهیم پایه قرانی

	لیدز	گراف پایه نهایی
تعداد مفاهیم موجود	۶۵	۲۵۵۷۹
تعداد انواع روابط بین نمونه‌ها	۷ نوع	۵۰ نوع
تعداد انواع روابط بین مفاهیم	۶ نوع	۱۰۶ نوع
تعداد نمونه‌ها	۲۲۰	۹۶۶۵
تعداد روابط بین نمونه‌ها	۳۰	۲۴۸۵۵
تعداد روابط بین مفاهیم	۰	۴۳۷۱۷۷

جهت غنی‌سازی گراف مفاهیم قرانی ساخته شده سعی نمودیم منابع دانش مکملی را به آن اضافه و یا نگاشت نماییم. لذا از فارس‌نت و همچنین اصطلاحنامه علوم قرآنی جهت نگاشت مفاهیم موجود در گراف پایه استفاده گردید تا ارتباط تمامی اجزاء دانش برقرار گردد. همچنین در فرآیند غنی‌سازی، آرگومان‌های افعال تولید شده از ترجمه قرآن در قالب SRL به مفاهیم موجود در گراف مفاهیم قرآنی پایه نگاشت می‌شوند. در صورتیکه گراف مفاهیم فاقد مفهوم مورد نظر باشد، مفهوم جدید به گراف مفاهیم اضافه می‌شود. خود افعال SRL نیز مستقیما به فرهنگ لغت فارس‌نت نگاشت می‌شوند تا ابهام‌زدایی[۳۵] افعال جملات ترجمه قرآن انجام شود شکل ۳، روش تلفیق این منابع را برای جمله «و تو (موسی) (در مصر) یکی (از فرعونیان) را کشتی؛ اما ما (خدا) تو را ... نجات دادیم» (آیه ۴۰ سوره طه) نشان می‌دهد.

روند نگاشت به صورت زیر است:

۱- جهت نگاشت از عناصر گراف مفهومی قرانی به سمت واژگان فارس‌نت بوده و به صورت یک اشاره‌گر به SYNSet مرتبط با آن عنصر درون فارس‌نت باشند. علت این که جهت نگاشت از عناصر گراف مفهومی قرانی به سمت واژگان فارس‌نت انتخاب شده است، این بوده که نیازی وجود ندارد که همه عناصر درون فارس‌نت به درون گراف مفهومی قرانی آورده شوند بلکه نیاز است تا از فارس‌نت برای بسط و تعریف عناصر درون گراف مفهومی قرانی استفاده شود. با این تصمیم از بزرگ شدن بی‌جهت گراف مفهومی قرانی جلوگیری می‌شود.

۲- جهت نگاشت از عناصر گراف مفهومی قرانی به سمت عناصر اصطلاحنامه بوده و به صورت یک اشاره‌گر به مدخل مرتبط با آن عنصر درون اصطلاحنامه باشند. مجددا علت این که جهت نگاشت از عناصر گراف مفهومی قرانی به سمت عناصر اصطلاحنامه انتخاب شده است، جلوگیری از بزرگ شدن بی‌جهت گراف مفهومی قرانی بوده است.

۳- جهت نگاشت از آرگومان‌های افعالSRL به سمت عناصر گراف مفهومی قرانی بوده و به صورت یک اشاره‌گر به مفاهیم مرتبط با آن عنصر درون گراف مفهومی قرانی باشند. علت این که جهت نگاشت از عناصر SRL به سمت عناصر گراف مفهومی قرانی انتخاب شده است، این بوده که نیازی وجود ندارد که همه عناصر درون SRL به درون گراف مفهومی قرانی آورده شوند بلکه در فرایند نگاشت آنهایی که مورد نیاز باشند به گراف مفهومی قرانی اضافه می‌گردند و به عبارت دیگر گراف مفهومی قرانی به کمک عناصر SRL تکمیل می‌شوند. در این حالت فقط برای کاربردهای استنتاجی، نگاشتی از طرف عناصر SRL به سمت عناصر گراف مفهومی قرانی بر قرار می‌شوند. با این تصمیم از بزرگ شدن بی‌جهت گراف مفهومی قرانی نیز جلوگیری می‌شود.

بنابراین روند کلی بدین صورت خواهد بود که با اضافه شدن هر عنصر به گراف مفهومی قرانی، نگاشت‌های متناظر برای آن به منابع اصطلاحنامه، فارس‌نت، و SRL برقرار خواهد شد. این کار کمک خواهد کرد تا بتوان پاسخ‌هایی مرتبط‌تر به پرس‌وجوها برگرداند.

۵- فرایندهای سامانه

مطابق چارچوب پیشنهادی سامانه پرسش و پاسخ قرانی که در شکل ۲ آمده است، بطور کلی سامانه از سه بخش اصلی تشکیل شده است که عبارتند از:

الف) پردازش پرسش: برای پردازش پرسش سه رویکرد در نظر گرفته شد و پیاده‌سازی گردید:

۱- رویکرد لغوی: استخراج نوع پاسخ، هدف و همبافت: برای این کار، حدود ۱۰۰۰۰ پرسش قرانی (که شامل حدود ۶۰۰۰ پرسش از FAQهای قرانی+۳۰۰۰ پرسش تهیه شده توسط کارشناسان قرانی+۱۰۰۰ پرسش مربوط به معماهای قرانی) مورد بررسی قرار گرفت و یک درختواره برای پرسش‌های قرانی با ۷۵ نوع پاسخ مجزا شکل گرفت. پس از آن برای تمام پرسش‌های مذکور بصورت دستی برچسب نوع پاسخ، هدف و همبافت مشخص گردید تا به عنوان داده آموزشی مولفه‌های توسعه داده شده این بخش باشد.

۲- رویکرد نحوی: بطور کلی گراف مفاهیم شامل حدود ۱۵۶ نوع رابطه است که تنها ۷۵ تای آن با انوع پاسخ مشخص شده در درختواره پرسش مطابقت دارد لذا مابقی دانش گراف با فقط توسعه مولفه‌های بند بالا مورد استفاده قرار نخواهد گرفت. لذا در این راستا مولفه POV توسعه داده شد تا بتوانیم از تمام دانش نهفته در گراف برای پاسخدهی به پرسش‌های کاربران استفاده نماییم.

۳- رویکرد معنایی: با استفاده از موتور SRL توسعه داده شده در پروژه می‌توان پرسش‌ها را بصورت معنایی تجزیه کرد و با تعیین آرگومان‌های مربوط به فعل پرسش، مورد مجهول را مشخص نموده و در دانش به دنبال آن گشت.

ب) استخراج پاسخ: در این بخش سعی شد تا مولفه‌های متعددی با رویکردها و روش‌های مختلف پیاده‌سازی شوند تا بتوانیم علاوه بر پوشش ۷۵ نوع پاسخ مشخص شده در درختواره پرسش، به دقت خوبی در استخراج پاسخ‌ها دست یابیم. این مولفه‌ها از نظر کارکرد به دو بخش تقسیم می‌شوند: مولفه‌های مبتنی بر متن که از روش‌های مختلفی از قبیل داده‌کاوی، رخداد کلمات و غیره برای یافتن پاسخ استفاده می‌کنند و مولفه(ها) مبتنی بر معنا که از روش استدلال برای انتخاب پاسخ مناسب استفاده می‌نماید. در ادامه هر یک از این مولفه‌ها به اختصار توضیح داده می‌شوند:

- مولفه استخراج پاسخ مبتنی بر موجودیت‌های اسمی: این مولفه ۱۴ نوع از ۷۵ نوع پاسخ را که جواب آنها یک موجودیت اسمی است پاسخ می‌دهد. این مولفه مبتنی بر لیست عمل می‌کند؛ یعنی به ازای تمام این ۱۴ نوع پاسخ یک لیست مجزا که شامل نهادهای اسمی مربوطه است تشکیل می‌شود و پس از اینکه پرسش نوع پاسخش مشخص شد و مستندات مربوطه از لوسین بازیابی شدند، آنگاه تمامی موجودیت‌های اسمی مرتبط با نوع پاسخ از متن انتخاب شده و با الگوریتم خاصی پاسخ‌های کاندید رتبه‌بندی می‌گردند و در نهایت پاسخ(های) با امتیاز بالا به کاربر نشان داده می‌شود.
- مولفه استخراج پاسخ مبتنی بر آیه‌یاب: این مولفه علاوه بر اینکه ۳ نوع پاسخ از ۷۵ نوع را پوشش می‌دهد قابلیت‌های زیر را نیز دارا می‌باشد:
  - ابزاری برای تشخیص ارجاع به متن فارسی
  - به عنوان یک موتور جستجوی قرانی
  - یک سامانه پرسش و پاسخ محدود
- مولفه استخراج پاسخ مبتنی بر جدول: این مولفه به کمک جداول از قبل تهیه شده و با توجه به هدف و نوع پاسخ هر پرسش، مسئول پاسخ‌دهی به ۳ نوع پاسخ از ۷۵ نوع می‌باشد.
- مولفه استخراج پاسخ مبتنی بر گراف مفاهیم: این مولفه از برچسب‌های پرسش استفاده کرده و پرس‌وجوهای مناسبی به زبان SPARQL را شبیه‌سازی می‌نماید تا بتواند داخل گراف را بازیابی کرده و پاسخ مناسب را برگرداند. مولفه مذکور قابلیت پاسخگویی به ۸ نوع پاسخ از ۷۵ نوع را دارد.
- مولفه استخراج پاسخ مبتنی بر استدلال: این مولفه با استفاده از الگوهای استدلال تعریف شده در موتور آن و با ترکیب این الگوها قابلیت پاسخگویی به ۶۴ نوع پاسخ از بین ۷۵ نوع را دارد. منبع دانش این مولفه گراف مفاهیم قرانی است.
- مولفه POV: این مولفه از رویکرد نحوی (رویکرد دوم پردازش پرسش) استفاده کرده و پاسخ‌ها را از درون گراف استخراج می‌نماید.
- مولفه استخراج پاسخ مبتنی بر بازنمایی ترکیبی روابط (بِتِر): این مولفه از رویکرد معنایی (رویکرد سوم پردازش پرسش) استفاده کرده و به پرسش کاربر پاسخ می‌دهد. این روش یکی از سخت‌ترین فعالیت‌هایی است که استارت آن در این پروژه زده شد و امید است بتوان با سرمایه‌گذاری بیشتر روش پیشنهادی بِتِر را توسعه داد.

شاید بتوان گفت که با توسعه این روش به نحوی هوش مصنوعی محقق شده است و ماشین درک معناداری از داده¬ها خواهد داشت.

ج) ترکیب پاسخ: از آنجا که مولفه‌های متعددی برای پاسخگویی وجود دارند و به ازای برخی از انواع پاسخ، بیش از یک مولفه وظیفه پاسخگویی به پرسش(های) مربوطه را دارد، لذا مسئله ترکیب پاسخ‌ها از اهمیت ویژه‌ای برخوردار است. این مولفه علاوه بر اینکه پاسخ‌های بازگشتی یک مولفه را با هم ترکیب می‌کند، وظیفه ترکیب پاسخ‌های بازگشتی مولفه‌های مختلف را نیز برعهده دارد.

جدول۳: ارزیابی مولفه های استخراج پرسش با نوع پاسخ صحیح

با در نظر گرفتن نوع پاسخ صحیح			
	C@1	C@2	C@3
مولفه استخراج پاسخ مبتنی بر NER	٪۴۸	٪۵۳	٪۵۴
مولفه استخراج پاسخ مبتنی بر ایه یاب	٪۹۱	٪۹۱	٪۹۱
مولفه استخراج پاسخ مبتنی بر جدول	٪۹۵	٪۹۵	٪۹۵
مولفه استخراج پاسخ مبتنی بر گراف مفاهیم	٪۳۰	٪۳۰	٪۳۰
مولفه استخراج پاسخ مبتنی بر استدلال	٪۶۳	٪۷۰	٪۷۲
مولفه ترکیب پاسخ	٪۶۵	٪۷۴	٪۷۶

۶- تست و تحلیل نتایج

همانطور که در شکل۲مشخص است، مولفه‌های متعددی در طول اجرای پروژه ایجاد و توسعه یافتند که مورد تست و ارزیابی قرار گرفتند. داده آزمایشی جامع که شامل ۸۶۶ پرسش قرانی تهیه شده توسط کارشناسان قرانی بود بخش‌های مختلف مورد محک قرار گرفت. دقت کلیه مولفه‌های سامانه و نیز دقت کل سامانه در ادامه آمده است:

الف) مولفه‌های پردازش پرسش:

رویکردی اصلی برای پردازش نوع پرسش، رویکرد اول می باشد که مبتنی بر یادگیری ماشین است (Sherkat 2014). در رویکرد اول مهمترین مولفه، تعیین نوع پاسخ است که با یافتن ان سایر مولفه های سامانه سایر مراحل استخراج پاسخ را طی می نمایند. جدول ۲ نتایج ارزیابی مولفه های پردازش پرسش را نشان می دهد.

جدول۲: ارزیابی مولفه های پردازش پرسش

رویکرد اول		
مولفه استخراج نوع پاسخ		۷۵.۹۲
مولفه استخراج هدف		۷۷
مولفه استخراج همبافت		۹۵.۴۸
رویکرد دوم		
تجزیه نحوی	دادگان اصلی	۸۴.۷۶
	دادگان نو	۹۲.۸۳
رویکرد سوم		
تجزیه معنایی (موتور SRL)		۶۱

ب) مولفه‌های استخراج پاسخ:

در این سامانه از روش های مختلفی برای استخراج پاسخ صورت پذیرفته است. بسیاری از این روش ها مستقل از یکدیگر می باشند و تنها برای پاسخ دادن به یک یا چند نوع خاصی از پرسش ها (طبق درختواره ۷۵ تایی پرسش های سامانه) می باشند. در تعداد کمی از انواع ممکن است دو روش همزمان پاسخ پرسش را استخراج نمایند که بر اساس میزان کانفیدنس هر مولفه امتیاز نهایی پاسخ بدست می اید. نتایج گزارش شده برای مولفه ترکیب پاسخ، در حقیقت دقت نهایی سامانه با در نظر گرفتن تمامی انواع پاسخ می باشد و در حقیقت دقت نهایی سامانه است. دقت گزارش شده برای مولفه ها، در حقیقت دقت ان مولفه برای طیف خاصی از انواع پرسش ها می باشد و نه لزوما برای تمامی انواع پرسش ها است. از انجایی که مولفه نوع پاسخ تاثیر بسیاری در نتایج نهایی سامانه دارد، نتایج در دو حالت، با در نظر گرفتن نوع پاسخ صحیح (یعنی مولفه ی تعیین نوع پاسخ به درستی نوع پاسخ پرسش را مشخص نموده است) و در حالت درنظر گرفتن پاسخ صحیح و غیر صحیح این مولفه مورد بررسی قرار گرفته است. با بررسی نتایج جدول ۳ و ۴ به تاثیر بسیار مهم مولفه تعیین نوع پاسخ در نتایج سایر مولفه ها می توان پی برد.

ج) تست کل سامانه:

جدول های ۵ و ۶ به ترتیب نتایج نهایی سامانه را در حالت درنظر گرفتن نوع پاسخ صحیح و در حالت عدم در نظر گرفتن نوع پاسخ صحیح نشان می دهند. برای ارزیابی نهایی سامانه از دو نفر کارشناس قرآنی خارج از تیم توسعه سامانه کمک گرفته شده است. شرح کامل تر نتایج و ارزیابی های بیشتر سامانه در سایت سامانه قرار دارد.

جدول۴: ارزیابی مولفه های استخراج پرسش با نوع پاسخ صحیح و غیر صحیح

	C@1	C@2	C@3
مولفه استخراج پاسخ مبتنی بر NER	٪۳۱	٪۳۵	٪۳۶
مولفه استخراج پاسخ مبتنی بر ایه یاب	٪۸۳	٪۸۳	٪۸۳
مولفه استخراج پاسخ مبتنی بر جدول	٪۹۵	٪۹۵	٪۹۵
مولفه استخراج پاسخ مبتنی بر گراف مفاهیم	٪۲۷	٪۲۷	٪۲۷
مولفه استخراج پاسخ مبتنی بر استدلال	٪۵۵	٪۶۰	٪۶۳
مولفه ترکیب پاسخ	٪۵۶	٪۶۴	٪۶۷

جدول۵: نتایج نهایی سامانه با نوع پاسخ صحیح

نوع پاسخ صحیح	C@1	C@2	C@3
کل سامانه	۶۵.۱۰	۷۳.۶۶	۷۶.۳۴

جدول ۶: نتایج نهایی سامانه با نوع پاسخصحیح و نا صحیح

نوع پاسخ صحیح و ناصحیح	C@1	C@2	C@3
کل سامانه	۵۶.۴۲	۶۳.۹۲	۶۷.۰۷

۴-نتیجه‌گیری

در این مقاله به معرفی سامانه پرسش و پاسخ خودکار قرانی "قران‌جوی" پرداخته شد. این سامانه برای پاسخ به سوالات قرانی کاربران طراحی شده است. همانطور که در مقاله نیز ذکر شد، در این سامانه از روش‌های مختلف و متعددی جهت استخراج پاسخ استفاده شد تا حتی‌الامکان بتوان پاسخ دقیق را از منابع دانش سامانه استخراج نمود. یکی از مهم‌ترین منابع دانش سامانه، گراف مفاهیم قرانی توسعه داده شده بنام "قران‌نگار" است که تاثیر بسزایی در استخراج پاسخ‌ها داشته است. لازم به ذکر است معماری پیشنهادی می‌تواند برای هر سامانه پرسش و پاسخ دامنه بسته استفاده گردد. در ادامه قصد داریم با تقویت بخش معنایی سامانه به نتایج بهتری دست یابیم.

مراجع

[1] Baeza R. Yates and Ribeiro-Neto B., 1999, “Modern Information Retrieval”, ACM Press/Addison-Wesley.
[2] Kangavari, M., S. Ghandchi, and M. Golpour. 2008. “A New Model for Question Answering Systems.” In Proceedings of World Academy of Science, Engineering and Technology, 32:536–543.

[3] Kwok, C., O. Etzioni, and D. S. Weld. 2001. "Scaling Question Answering to the Web." ACM Transactions on Information Systems (TOIS) 19 (3): 242–262.
[4] Lampert, A., 2004. "A Quick Introduction to Question Answering." CSIRO ICT Centre.
[5] Lopez, V., M. Pasin, and E. Motta. 2005a. "Aqualog: An Ontology-portable Question Answering System for the Semantic Web." The Semantic Web: Research and Applications: 135–166.
[6] Schlaefer, N., P. Gieselmann, T. Schaaf, and A. Waibel. 2006. "A Pattern Learning Approach to Question Answering Within the Ephyra Framework." In Text, Speech and Dialogue, 687–694.
[7] Vargas-Vera, M., and E. Motta. 2004. "AQUA–ontology-based Question Answering System." MICAI 2004: Advances in Artificial Intelligence: 468–477.
[8] Vicedo, J., R. Izquierdo, F. Llopis, and R. Munoz. 2004. "Question Answering in Spanish." Comparative Evaluation of Multilingual Information Access Systems: 27–38.
[9] Weizenbaum, J. 1966. "ELIZA—a Computer Program for the Study of Natural Language Communication Between Man and Machine." Communications of the ACM 9 (1): 36–45.
[10] Zheng, Z. 2002. "AnswerBus Question Answering System." In Proceedings of the Second International Conference on Human Language Technology Research, 399–404
[11] Unger, Christina, et al. "Template-based question answering over RDF data." Proceedings of the 21st international conference on World Wide Web. ACM, 2012.
[12] Dwivedi, Sanjay K., and Vaishali Singh. "Research and reviews in question answering system." Procedia Technology 10 (2013): 417-424.
[13] Sherkat, Ehsan, and Mojgan Farhoodi. "A hybrid approach for question classification in Persian automatic question answering systems." Computer and Knowledge Engineering (ICCKE), 2014 4th International eConference on. IEEE, 2014.

[۱۴] حجازی, محمد رضا, مریم سادات میریان حسین آبادی, کورش نشاطیان, بهادر رضا افقی، احسان درودی. ۱۳۸۳. "سامانه پرسش و پاسخ برای حوزه مخابرات با قابلیت استخراج و دسته‌بندی خودکار مستندات." علوم و مهندسی کامپیوتر نشریه علمی پژوهشی انجمن کامپیوتر ایران.
[۱۵] مژگان فرهودی و شهریار سموری. ۱۳۸۹. "مدل مفهومی سامانه پرسش و پاسخ خودکار". مرکز تحقیقات مخابرات ایران.

زیرنویس

[1] Interactivity
[2] Scalable
[3] http://quranjooy.itrc.ac.ir/
[4] IBM
[5] Yahoo
[6] ELIZA
[7] MIT
[8] DOCTOR
[9] LUNAR
[10] Touring Test
[11] Open domain
[12] Framework
[13] Alicante
[14] Portable
[15] Ontology
[16] Semantic Web
[17] Real time
[18] Watson
[19] Natural Language Processing (NLP)
[20] Lampert
[21] Question processing
[22] Question/Answering Classification
[23] Question Focus
[24] Keyword Extraction
[25] Paragraph Indexing
[26] Paragraph Filtering
[27] Paragraph Quality
[28] Paragraph Ordering
[29] Answer Processing
[30] Answer Extraction
[31] Answer Ranking
[32] Answer Correctness
[33] Reasoning Engine
[34] http://corpus.quran.com/ontology.jsp
[35] Disambiguation