

ساخت یک سیستم پرسش و پاسخ برای کیوسک اطلاعات دانشکده مهندسی برق و کامپیوتر

نیره کرمانشاهی^۱، حسین عبدالکریمی^۲، مهرنوش شمس فرد^۳

^۱ آزمایشگاه پردازش زبان طبیعی، دانشکده مهندسی برق و کامپیوتر، دانشگاه شهید بهشتی، تهران، ایران،
n.kermanshahi@mail.sbu.ac.ir

^۲ آزمایشگاه پردازش زبان طبیعی، دانشکده مهندسی برق و کامپیوتر، دانشگاه شهید بهشتی، تهران، ایران،
h.abdolkarimi@mail.sbu.ac.ir

^۳ استادیار، گروه مهندسی کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه شهید بهشتی، تهران، ایران
m-shams@sbu.ac.ir

چکیده

در این مقاله یک سیستم پرسش و پاسخ به زبان فارسی برای دامنه محدود کیوسک اطلاعات دانشکده مهندسی برق و کامپیوتر معرفی می‌شود. محدوده سؤالات ورودی در این سیستم شامل انواع سؤالاتی است که معمولاً در دانشکده از اطلاعات یا آموزش پرسیده می‌شود. سیستم قابلیت پاسخگویی به سؤالات مکانی، زمانی، اشخاص، دروس و ویژگی‌های این موجودیت‌ها را خواهد داشت. رویکرد پیشنهادی در این سیستم یک رویکرد معنایی مبتنی بر استفاده از هستان‌شناسی و بازنمایی پرسش به صورت سه تایی‌های RDF است. در این سیستم پرسش کاربر با استفاده از هستان‌شناسی سیستم به مجموعه‌ای از سه تایی‌های RDF (فاعل، فعل، مفعول) تبدیل می‌گردد. کلید استخراج پاسخ، نگاشت مناسب بین سه تایی‌های پرسش و روابط موجود در هستان‌شناسی است. به‌کارگیری رویکردهای معنایی و مبتنی بر دانش منجر به تولید سیستم پرسش - پاسخی با دقت پاسخگویی ۹۱.۳۴٪ و فراخوان ۸۷.۹۲٪ گردیده است.

کلمات کلیدی

سیستم پرسش و پاسخ، هستان‌شناسی، سه تایی‌های RDF، تحلیل و بسط پرسش، استخراج و تولید پاسخ، پردازش زبان فارسی.

منابع در دسترسشان استخراج می‌نمایند. در این قسمت مروری
اجمالی بر برخی سیستم‌های پرسش و پاسخ با توجه به نوع بردار
ویژگی پرسش و نوع منابع پاسخ، با تاکید بر ساختار و قابلیت نگاشت
این دو انجام می‌دهیم.

در بین سیستم‌های پرسش و پاسخ مورد بررسی، برخی هم در
بردار ویژگی و هم در منابع دانش خود فاقد ساختار و ارتباطی قوی
بین اجزا می‌باشند. بردار ویژگی شامل کلمات کلیدی و مترادف‌های
آن‌ها، ریشه کلمات و موارد دیگری با محوریت کلمه می‌باشد اما ارتباط
بین کلمات پرسش نادیده گرفته می‌شود. منابع دانش نیز منابع غیر
ساخت یافته متنی یا وب می‌باشد. از این دسته می‌توان سیستم
AnswerBus [۴] را نام برد.

۱- مقدمه

پرسش و پاسخ یکی از موضوعات باز در حوزه پردازش زبان‌های طبیعی
است که به ارائه پاسخ کوتاه و دقیق برای پرسش‌های به زبان طبیعی
می‌پردازد. در سال‌های اخیر پژوهش‌های متعددی در زمینه ساخت
سیستم‌های پرسش و پاسخ با دامنه باز یا محدود صورت گرفته است.
این سیستم‌ها دارای شباهت‌ها و تفاوت‌هایی هستند. برخی وابسته به
زبانند و به راحتی برای زبان‌های دیگر قابل به‌کارگیری نمی‌باشند
درحالی‌که برخی دیگر از تکنیک‌های مستقل از زبان استفاده می‌کنند.
عمده این سیستم‌ها ابتدا پرسش ورودی را به یک بازنمایی میانی
(مانند بردار ویژگی) تبدیل می‌کنند و سپس بر اساس آن پاسخ را از