

یادگیری هستان‌شناسی از پایگاه داده رابطه‌ای

بتول لکزائی، مهرنوش شمس‌فرد

آزمایشگاه پردازش زبان طبیعی، دانشکده مهندسی برق و کامپیوتر، دانشگاه شهید بهشتی، تهران، ایران،

b.lakzaei@yahoo.com, m-shams@sbu.ac.ir

چکیده

هستان‌شناسی یک توصیف صریح و صوری از یک مفهوم‌سازی اشتراکی است. ساخت دستی هستان‌شناسی‌ها امری بسیار پرهزینه، وقت‌گیر و مستعد خطا است. راه حل رفع این مشکلات خودکار سازی فرآیند ساخت هستان‌شناسی یا به عبارت دیگر یادگیری هستان‌شناسی از منابع مختلف از جمله منابع ساخت‌یافته مانند پایگاه‌های داده است.

امروزه پایگاه‌های داده رابطه‌ای، یکی از بهترین روش‌ها برای ذخیره و بازیابی اطلاعات محسوب می‌شوند اما برای اهداف مختلفی که مبتنی بر نمایش دانش معنایی می‌باشند، مناسب نیستند. از این رو، در این مقاله روشی نوین برای استخراج خودکار یک هستان‌شناسی از شمای یک پایگاه داده رابطه‌ای پیشنهاد شده است. در مقایسه با سایر روش‌های موجود، این روش علاوه بر تحلیل جدول‌ها، کلیدهای اصلی و خارجی و داده‌های موجود در چندتایی‌ها، با تحلیل محدودیت‌های تعریف شده روی ستون‌های یک جدول و نیز تحلیل دیگر عناصر موجود در شمای رابطه‌ای مانند دید، توابع تعریف شده توسط کاربر و رویه‌های ذخیره شده، به استخراج خودکار عناصر هستان‌شناسی می‌پردازد. در نهایت این روش، با یک هستان‌شناسی مرجع که به طور دستی در قلمرو مورد نظر ساخته شده است، مورد مقایسه قرار گرفته و معیارهای فراخوان، دقت و معیار F- به منظور ارزیابی کیفیت هستان‌شناسی تولید شده، محاسبه شده است. مقایسه این معیارها، بیانگر این است که روش پیشنهادی، نسبت به سایر روش‌های موجود عملکرد بهتری دارد.

کلمات کلیدی

هستان‌شناسی، یادگیری هستان‌شناسی، پایگاه داده رابطه‌ای

گروبر [2] هستان‌شناسی را به عنوان یک توصیف صریح و صوری از یک مفهوم‌سازی اشتراکی، تعریف می‌کند. در این تعریف، "مفهوم‌سازی" به معنای یک مدل انتزاعی از پدیده‌های جهان و مفاهیم مرتبط با این پدیده‌ها است. منظور از "صریح" این است که مفاهیم مورد استفاده در این مدل و محدودیت‌های آن‌ها به طور صریح تعریف می‌شوند. "صوری بودن" این توصیف به معنای مدل‌پذیری هستان‌شناسی با هدف قابل فهم بودن آن برای ماشین می‌باشد. در نهایت، منظور از "اشتراکی" بودن، این است که هستان‌شناسی باید شامل دانش مشترکی باشد که مورد پذیرش جوامع علمی مختلف می‌باشد.

۱- مقدمه

وب جهان‌گستر^۱، یک منبع پهناور و رو به رشد از اطلاعات و سرویس‌هایی است که توسط افراد و برنامه‌های کاربردی مختلف به اشتراک گذاشته می‌شوند [1]. در واقع، وب در حال تکامل از یک فضای اطلاعاتی و ارتباطاتی گسترده به سمت مخزنی انبوه از دانش و سرویس‌های مختلف می‌باشد و هستان‌شناسی^۲ یکی از عواملی است که امکان چنین تکاملی را فراهم می‌کند [2]. هستان‌شناسی نقش مهمی در پشتیبانی از تبادل و به اشتراک‌گذاری اطلاعات، از طریق فراهم کردن امکان تعامل‌پذیری^۳ معنایی در وب، ایفا می‌کند.