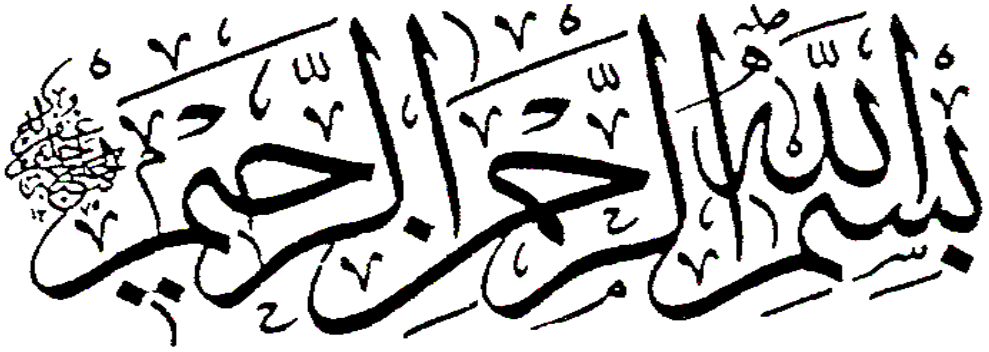

بسم الله الرحمن الرحيم

**دانشکده آموزش های الکترونیکی**

**پایان نامه کارشناسی ارشد در رشته مهندسی فناوری اطلاعات (تجارت الکترونیک)**

# ارائه الگویی جهت پیاده سازی جستجوی معنایی با استفاده از تکنیکهای وب معنایی

**توسط**

**امیر هراتی علمداری**

**استاد راهنما:**

**دکتر امید بوشهریان**

**مهرماه ۱۳۹۱**

# اظهارنامه

**اینجانب امیر هراتی علمداری دانشجوی رشته ی مهندسی فناوری اطلاعات گرایش تجارت الکترونیک دانشکده ی آموزشهای الکترونیکی اظهار می کنم که این پایان نامه حاصل پژوهش خودم بوده و در جاهایی که از منابع دیگران استفاده کرده ام، نشانی دقیق و مشخصات کامل آن را نوشته ام. همچنین اظهار می کنم که تحقیق و موضوع پایان نامه ام تکراری نیست و تعهد می نمایم که بدون مجوز دانشگاه دستاوردهای آن را منتشر ننموده و یا در اختیار غیر قرار ندهم. کلیه حقوق این اثر مطابق با آیین نامه مالکیت فکری و معنوی متعلق به دانشگاه شیراز است.**

**نام و نام خانوادکی : امیر هراتی علمداری**

**تاریخ و امضا :**

# تقدیم به

## پدر و مادر عزیزم

خدای را بسی شاکرم که از روی کرم پدر و مادری فداکار نصیبم ساخته تا در سایه درخت پر بار وجودشان بیاسایم و از ریشه آنها شاخ و برگ گیرم و از سایه وجودشان در راه کسب علم ودانش تلاش نمایم. والدینی که بودنشان تاج افتخاری است بر سرم و نامشان دلیلی است بر بودنم چرا که این دو وجود پس از پروردگار مایه هستی ام بوده اند دستم را گرفتند و راه رفتن را در این وادی زندگی پر از فراز و نشیب آموختند. آموزگارانی که برایم زندگی؛ بودن و انسان بودن را معنا کردند.

حال این برگ سبزی است تحفه درویش تقدیم آنان...

به پاس تعبیر عظیم و انسانی شان از کلمه ایثار و از خودگذشتگان، به پاس عاطفه سرشار و گرمای امیدبخش وجودشان که در این سردترین روزگاران بهترین پشتیبان است، به پاس قلب های بزرگشان که فریاد رس است و سرگردانی و ترس در پناهشان به شجاعت می گراید و به پاس محبت های بی دریغشان که هرگز فروکش نمی کند.

# سپاسگزاری

سپاس خدای را که سخنوران، در ستودن او بمانند و شمارندگان، شمردن نعمت های او ندانند و کوشندگان، حق او را گزاردن نتوانند. اکنون که این رساله به پایان رسیده است شایسته است که از زحمات استاد فرهیخته و فرزانه **جناب آقای دکتر امید بوشهریان** که با کرامتی چون خورشید، سرزمین دل را روشنی بخشیدند و گلشن سرای علم و دانش را با راهنمایی های کار ساز و سازنده بارور ساختند؛ تقدیر و تشکر نمایم.

همچنین از پدر و مادر عزیز، دلسوز و مهربانم که آرامش روحی و آسایش فکری فراهم نمودند تا با حمایت های همه جانبه در محیطی مطلوب، مراتب تحصیلی و نیز پایان نامه درسی را به نحو احسن به اتمام برسانم؛ سپاسگزاری نمایم.

**نام : امیر** **نام خانوادگی: هراتی علمداری**

**مقطع تحصیلی: کارشناسی ارشد** **رشته وگرایش: مهندسی فناوری اطلاعات گرایش تجارت الکترونیک**

**استادراهنما: دکتر امید بوشهریان** **تاریخ دفاع: ۱۳۹۱/۷/۳**

# ارائه الگویی جهت پیاده سازی جستجوی معنایی

# با استفاده از تکنیک های وب معنایی

آنچه که تا کنون در فضای وب جریان داشته است افزایش روزافزون حجم اطلاعات و مستندات موجود در آن بوده است، بدون اینکه این افزایش حجم مستندات تاثیری در کاربرد و بهره برداری خودکار اطلاعات موجود به همراه داشته باشد. به علاوه امروزه اکثریت محتوای وب بر روی پایگاه داده های رابطه ای ذخیره شده اند.

حرکتی که گام های اولیه آن درحال شکل گیری است تغییر شکل «وب» به موجودیتی تحت عنوان «وب معنایی» میباشد. وب معنایی یک گسترش از وب موجود است که گاهی اوقات تحت عنوان «وب آینده» و یا «وب نسخه ۳» معرفی می شود، که در آن به اطلاعات، معانی خوش تعریفی داده می شود که بتوانند قابلیت همکاری بهتر بین کامپیوتر ها و انسان را فراهم کنند.

موفقیت وب معنایی بر تولید انبوه داده های معنایی وابسته است. با توجه به حجم عظیم داده های رابطه ای ذخیره شده در پایگاه داده های رابطه ای موجود در وب، انتشار این داده های رابطه ای بر روی وب معنایی بسیار پر کاربرد خواهد بود.

در این پایان نامه، الگویی جهت پیاده سازی جستجوی معنایی بر روی داده های ذخیره شده در پایگاه داده های رابطه ای، با استفاده از تکنیک های وب معنایی ارائه شده است. گام های اصلی الگوی پیشنهادی در این پایان نامه از این قرارند:

۱-بازنمایی داده های سازمانی ذخیره شده بر روی پایگاه داده های رابطه ای بر اساس گراف های معنایی با استاندارد RDF به صورت یک پایگاه دانش معنایی(نگاشت داده های رابطه ای به گراف های معنایی)

۲-آنالیز وابستگی های داده ای برنامه و بازنمایی اجزای برنامه و ارائه یک آنتولوژی سازمانی بر اساس آنها

۳-پیاده سازی یک موتور جستجوگر معنایی که بر پایه آنتولوژی سازمانی تهیه شده و پایگاه دانش نگاشت شده، که به جستجوی معنایی بر روی اطلاعات دامنه مورد نظر می پردازد.

# فهرست مطالب


**۱. فصل اول : مقدمه ..... ۱**

۱.۱. مقدمه ای بر وب معنایی ..... ۲

۱.۲. طرح مساله ..... ۵

۱.۳. اهداف تحقیق ..... ۷

۱.۴. روش انجام تحقیق و دستاوردها ..... ۸

*۱.۴.۱. مراحل تحقیق ..... ۸*

۱.۴.۲. محصول پایان نامه ..... ۹

۱.۵. ساختار پایان نامه ..... ۱۱

**۲. فصل دوم : پیش زمینه و تحقیقات پیشین ..... ۱۳**

۲.۱. اجزای اساسی وب معنایی ..... ۱۳

۲.۲. معماری وب معنایی ..... ۱۶

۲.۳. آشنایی با RDF(Resource description Framework) ..... ۱۹

*۲.۳.۱. مدل RDF ..... ۲۰*

۲.۴. آشنایی با RDFs(RDF Schema) ..... ۲۱

۲.۵. آنتالوژی ..... ۲۴

۲.۶. زبان OWL ..... ۲۵

۲.۷. نحو ایجاد آنتالوژی ..... ۲۶

۲.۸. زبان پرس وجوی SPARQL ..... ۲۷

*۲.۸.۱. ساختار بازجست در SPARQL ..... ۲۷*

۲.۹. تحقیقات پیشین ..... ۲۸

*۲.۹.۱. نگاشت پایگاه داده های رابطه ای به آنتولوژی ..... ۲۹*

*۲.۹.۲. مهندسی نرم افزار بر پایه وب معنایی ..... ۳۰*

*۲.۹.۳. مدل سازی اجزای نرم افزار بر پایه آنتولوژی ..... ۳۲*

*۲.۹.۴. پایگاه دانش DBPedia ..... ۳۵*

**۳. فصل سوم : راهنمای ساخت آنتولوژی ..... ۳۷**

۳.۱. مقدمه ..... ۳۷

۳.۲. آنتولوژی چیست؟ ..... ۳۷
۳.۳. یک متدولوژی ساده مهندسی دانش ..... ۳۹
*۳.۳.۱. گام اول :تعیین دامنه و محدوده آنتولوژی ..... ۴۰*
سوالات صلاحیت ..... ۴۱
*۳.۳.۲. گام دوم: در نظر گرفتن امکان استفاده مجدد از آنتولوژی ..... ۴۲*
*۳.۳.۳. گام سوم: شمارش لغات مهم آنتولوژی ..... ۴۳*
*۳.۳.۴. گام چهارم: تعریف کلاس‌ها و سلسله مراتب کلاس ..... ۴۳*
*۳.۳.۵. گام پنجم: تعریف خصوصیات کلاس‌ها-اسلات ها ..... ۴۶*
*۳.۳.۶. گام ششم: تعریف جنبه های (فست) اسلات ها ..... ۴۸*
*۳.۳.۷. گام هفتم: ساخت نمونه ها ..... ۵۳*

**۴. فصل چهارم: روش پیشنهادی درجستجوی معنایی درون سازمانی ..... ۵۴**

۴.۱. مقدمه ..... ۵۴
۴.۲. تحلیل نمونه کاربردی ..... ۵۵
*۴.۲.۱. کاربران سیستم ..... ۵۷*
*۴.۲.۲. فرم‌های سیستم ..... ۵۷*
۴.۳. آماده کردن پایگاه داده ..... ۶۰
۴.۴. روش پیشنهادی برای استخراج آنتولوژی های مورد نیاز ..... ۶۲
*۴.۴.۱. استخراج اجزای اصلی آنتولوژی ..... ۶۴*
*۴.۴.۲. ایجاد نگاشت ..... ۶۷*
*۴.۴.۳. استخراج آنتولوژی داده ای از پایگاه داده رابطه ای ..... ۶۹*
*۴.۴.۴. ارائه آنتولوژی سازمانی ..... ۷۳*
*۴.۴.۵. پیاده سازی جستجوگر معنایی ..... ۸۱*
۴.۵. گامهای بازنمایی دانش سازمانی- داده ای بر اساس مدل پیشنهادی ..... ۸۱

**۵. پیاده سازی جستجوگر معنایی درون سازمانی ..... ۸۳**

۵.۱. مقدمه ..... ۸۳
۵.۲. مشخصات تکنیکی جستجوگر معنایی طراحی شده ..... ۸۳
۵.۳. پرس و جو های SPARQL ..... ۸۴
*۵.۳.۱. پرس و جو های آنتولوژی سازمانی ..... ۸۷*
*۵.۳.۲. پرس و جو های آنتولوژی سازمانی ..... ۸۸*

۵.۴. بررسی شیوه عملکرد جستجوگر معنایی ۸۹
*۵.۴.۱. پیش نیاز های اجرای جستجوگر معنایی ۹۰*
*۵.۴.۲. مثالی از اجرای جستجوگر معنایی ۹۱*
**۶. مقایسه مدل پیشنهادی با مدل‌های موجود ۹۹**
۶.۱. مقایسه روش‌های نگاشت آنتولوژی از پایگاه داده رابطه ای ۹۹
۶.۲. مقایسه روش‌های مدلسازی اجزای برنامه بر پایه آنتولوژی ۱۰۴
**۷. نتیجه گیری و کارهای آینده ۱۰۸**
۷.۱. مزایای روش پیشنهادی ۱۰۹
**۸. مراجع ۱۱۲**
**۹. پیوست‌ها ۱۱۹**
۹.۱. پیوست اول: انتشار پایگاه داده رابطه ای بر روی وب معنایی ۱۱۹
*۹.۱.۱. مقدمه ۱۱۹*
*۹.۱.۲. D2R Server ۱۱۹*
*۹.۱.۳. مشخصات زبان D2RQ ۱۲۰*
*۹.۱.۴. ClassMap-ایجاد منابع RDF ۱۲۵*
*۹.۱.۵. Property Bridge-افزودن خصوصیات به منابع ۱۲۹*
*۹.۱.۶. نگاشت شرطی ۱۳۷*
*۹.۱.۷. جداول تبدیل ۱۳۹*
*۹.۱.۸. به کار گرفتن فرهنگ لغات کلاس‌ها وخصوصیات ۱۴۲*
۹.۲. پیوست دوم: بخش‌هایی از فایل نگاشت سفارشی ۱۴۶

# فهرست جداول

جدول ۱. نوع‌های داده XSD برای نگاشت از SQL ..... ۵۰
جدول ۲. خصوصیات کلاس User برای درج در نگاشت ..... ۶۹
جدول ۳. نتایج مقایسه روش‌های مختلف نگاشت RDB به RDF [۴۲] ..... ۱۰۳
جدول ۴. نتایج ارزیابی سرعت اجرای بازجست ها در D2RQ [۲۵] ..... ۱۰۴
جدول ۵. دسته بندی آنتولوژی های مرتبط با مهندسی نرم افزار [۶۲] ..... ۱۰۴
جدول ۶. مقایسه روش‌های موجود استخراج آنتولوژی از اجزا برنامه ..... ۱۰۷
جدول ۷. خصوصیات d2rq:Database ..... ۱۲۴
جدول ۸. خصوصیات d2rq:ClassMap ..... ۱۲۸
جدول ۹. خصوصیات d2rq:PropertyBridge ..... ۱۳۳
جدول ۱۰. خصوصیات d2rq:TranslationTable ..... ۱۴۰
جدول ۱۱. خصوصیات d2rq:Translation ..... ۱۴۰
جدول ۱۲. خصوصیات d2rq:AditionalProperty ..... ۱۴۴

# فهرست تصاویر


شکل ۱. ایده اولیه آقای برنرز-لی درباره وب[۱] ..... ۲
شکل ۲. وب کنونی در مقایسه با وب معنایی[۴] ..... ۴
شکل ۳. بلوک دیاگرام کلی طرح ..... ۱۰
شکل ۴. گراف RDF ..... ۱۴
شکل ۵. گراف RDF نمونه ..... ۱۵
شکل ۶. معماری پشته ای وب معنایی[۵] ..... ۱۶
شکل ۷.گراف RDF نمونه ..... ۲۱
شکل ۸. گراف یک نمونه از RDF SCHEMA ..... ۲۳
شکل ۹. متدولوژی ایجاد آنتولوژی[۱۸] ..... ۲۷
شکل ۱۰. مدلسازی اجزای نرم افزار بر پایه آنتولوژی [۷۸] ..... ۳۵
شکل ۱۱. بعضی کلاس‌ها، نمونه‌ها و روابط میان آن‌ها دردامنه WINE. ..... ۳۹
شکل ۱۲. سطوح مختلف سلسله مراتب کلاس WINE ..... ۴۵
شکل ۱۳. اسلات ها کلاس WINEو جنبه (FACET) های این اسلات ها ..... ۴۷
شکل ۱۴. تعریف اسلات PRODUCESاز کلاس WINERY ..... ۵۰
شکل ۱۵. تعریف یک نمونه از کلاس BEAUJOLAIS ..... ۵۳
شکل ۱۶. فرم شناسنامه شرکت‌های همکار بخش رتبه و رشته ..... ۵۸
شکل ۱۷. فرم شناسنامه شرکت‌های همکار بخش تعداد نیروی انسانی ..... ۵۸
شکل ۱۸. فرم شناسنامه شرکت‌های همکار بخش اطلاعات مالی ..... ۵۹
شکل ۱۹. گزارش جامع سیستم ..... ۶۰
شکل ۲۰. ساختار پایگاه داده رابطه ای نمونه کاربردی ..... ۶۱
شکل ۲۱. خصوصیات جداول پایگاه داده نمونه کاربردی ..... ۶۲
شکل ۲۲. خصوصیات کلاس USER ..... ۶۶
شکل ۲۳. معماری نمونه برای یک نگاشت D2RQ [۴۳] ..... ۶۸
شکل ۲۴. معماری D2R SERVER [۲۵] ..... ۷۱
شکل ۲۵. بخش‌هایی از گرافهای RDF آنتولوژی داده ای ..... ۷۲
شکل ۲۶. صفحه دانلود آنتولوژی و DATASETهای موجود پایگاه دانش DBPEDIA ..... ۷۴
شکل ۲۷. فرایند استخراج آنتولوژی از ساختاروارثت برنامه‌ها در پروژه TAO [۴۴] ..... ۷۵
شکل ۲۸. TAXONOMYآنتولوژی GUI [۷۸] ..... ۷۷

شکل ۲۹. TAXONOMY آنتولوژی سازمانی ۷۸
شکل ۳۰. بخشهایی از گراف آنتولوژی سازمانی ۷۹
شکل ۳۱. مقدار دهی دو زبانه برای خصوصیت TEXTنمونه های کلاس CONTROLS ۸۱
شکل ۳۲. نحوه اجرای بازجست های SPARQL ۸۵
شکل ۳۳ PREFIXهای مورد نیاز برای بازجست SPARQL ۸۷
شکل ۳۴. بازجست اسامی فرم‌ها ۸۷
شکل ۳۵. بازجست اسامی LABLEها ی فرم انتخاب شده ۸۷
شکل ۳۶. بازجست نام فیلد مرتبط با LABLE انتخاب شده ۸۸
شکل ۳۷. بازجست نام جدول مرتبط با فرم انتخاب شده ۸۸
شکل ۳۸. بازجست مقادیر غیر تکراری خصوصیت انتخاب شده ۸۸
شکل ۳۹. بازجست تعیین نوع منبع ۸۹
شکل ۴۰. جستجوگر معنایی پیاده سازی شده ۹۰
شکل ۴۱. اجرای موتور D2R ۹۱
شکل ۴۲. لیست فرم‌های نمونه کاربردی ۹۲
شکل ۴۳. لیست برچسب‌های فرم انتخاب شده ۹۳
شکل ۴۴. روابط مابین اجزا در بخشی از گراف آنتولوژی سازمانی ۹۴
شکل ۴۵. افزودن فیلترها به جستجو ۹۵
شکل ۴۶. نتایج جستجوی معنایی ۹۵
شکل ۴۷. جزئیات عنصر جستجو شده ۹۶
شکل ۴۸. امکانات اضافی جستجوگر معنایی ۹۷
شکل ۴۹. معماری D2R SERVER ۱۲۰
شکل ۵۰. معماری نگاشت D2RQ ۱۲۱

# ۱. فصل اول : مقدمه

بیش از چند دهه از پیدایش وب می‌گذرد. اقبال روز افزون این پدیده، محققان را به توسعه هر چه بیشتر آن فراخوانده است. شبکه وب به عنوان یکی از بزرگ‌ترین دستاورد های تکنولوژی، به طور قابل ملاحظه ای بر جنبه های مختلفی ازجمله اجتماعی، اقتصادی، فرهنگی، تاثیر گذاشته است، به گونه ای که امروزه داده‌ها و اطلاعات گوناگونی در اختیار همگان قرار گرفته است و عملا وب به یک کتابخانه بزرگ از منابع تبدیل شده است. اما با وجود این پیشرفت‌ها، وضعیت کنونی وب رضایت بخش نیست زیرا بسیاری از امور ضروری و روز مره توسط ابزار های خودکار پشتیبانی نمی‌شوند. برای مثال موتور های جستجوی موجود که بر اساس کلمات کلیدی کار می‌کنند و مهم‌ترین ابزار برای یافتن اطلاعات در سطح وب هستند، ولی هنوز هنگام استفاده از آن‌ها مشکلات و محدودیت‌های اساسی وجود دارد. ازجمله اینکه نتایج یافت شده بسیار زیاد بوده و ممکن است اغلب آن‌ها هیچ ربطی با عبارت مورد جستجوی ما نداشته باشند.

آنچه که تا کنون در فضای وب جریان داشته افزایش روز افزون حجم اطلاعات و مستندات موجود در آن بوده است، بدون اینکه این افزایش حجم مستندات تاثیری در کاربرد و بهره برداری خودکار اطلاعات موجود به همراه داشته باشد. به بیان دیگر مسئولیت پالایش و بازیابی اطلاعات مفید و قابل استفاده از میان حجم گسترده اطلاعات موجود همواره با خود کاربر بوده است. حرکتی که گامهای اولیه آن امروزه در حال پیدایش است تغییر شکل «وب» به موجودیتی دیگر به نام «وب معنایی» می‌باشد. اساس این حرکت نوین، هدفمند ساختن اطلاعات و مستندات موجود و معنا بخشیدن به محتویات وب به گونه ای است که توسط کامپیوتر قابل فهم و پردازش باشد. [۱]

وب معنایی را می‌توان به فضایی جهانی از جنس محاسبات هوشمند ماشینی تصوّر کرد که در آن تمامی کتاب‌ها، کتاب‌خانه‌ها دانش‌ها، دانش‌نامه‌ها و دانشگاه‌ها و پایگاه‌های دانش[1] به صورتی معنی‌گرا و با توانایی درک مفهومی همدیگر، در کنار هم قرار می‌گیرند.

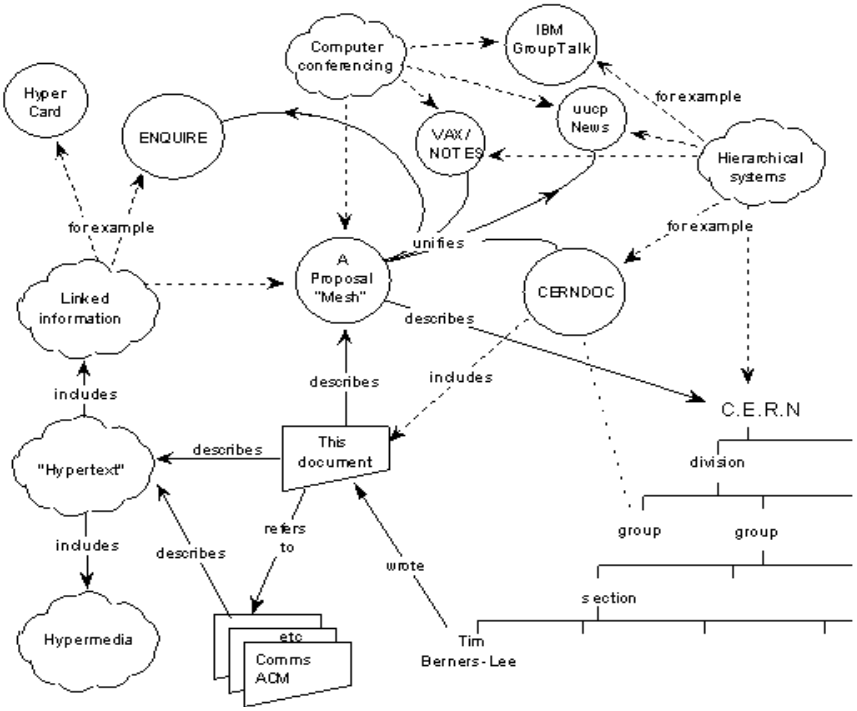
[1] Knowledge bases

آقای تیم برنرزلی[۲]، مخترع وب، در به تصویر کشیدن چشم انداز خود برای ابداعش اینگونه اظهار می‌کند که «هنگامی که وب معنایی ظهور کند، ساز و کارهای روزمره تجارت، بروکراسی و زندگی ما توسط ماشین‌هایی انجام خواهد شد که با یکدیگر صحبت می‌کنند...»[۲].

در این بخش جهت آشنایی با محور های اصلی این تحقیق، به معرفی مقدماتی از وب معنایی و همچنین طرح مساله، هدف و روش انجام، دستاوردها و ساختار کلی مطالب این پایان نامه خواهیم پرداخت.

## ۱.۱. مقدمه ای بر وب معنایی

ایده اولیه آقای برنرزلی درباره اینترنت با آنچه که ما امروزه از اینترنت می شناسیم، متفاوت است. از نظر ایشان وب آینده شامل دو بخش مهم است: یکی امکان همکاری بیشتر گروه ها و رسانه ها و دیگری ایجاد یک شبکه وب قابل فهم و قابل پردازش توسط ماشین [۳].




شکل ۱. ایده اولیه آقای برنرزلی درباره وب[۱]

[2] Tim Berners-Lee

شکل شماره ۱ ایده اولیه آقای برنرزلی را نشان می دهد [۱] . همانطور که دیده مـی شـود، از نظر ایشان، وب چیزی فراتر از مجموعه ای از صفحات HTML می باشد و باید امکـان تعریـف روابـط بـین منابع مختلـف موجـود در وب، وجـود داشـته باشـد. ماننـد روابـط describes، includes، wrote و غیـره. متاسفانه در وب کنونی، این تعاریف دیده نمی شوند. در واقع در ایده اولیه، عـلاوه بـر اطلاعـاتی کـه بـرای استفاده انسان ها بر روی وب قرار می گرفت، مقداری متادیتا[3] نیز برروی آنها قـرارداده مـی‌شـد تـا ایـن اطلاعات توسط ماشین ها قابل پردازش باشند.

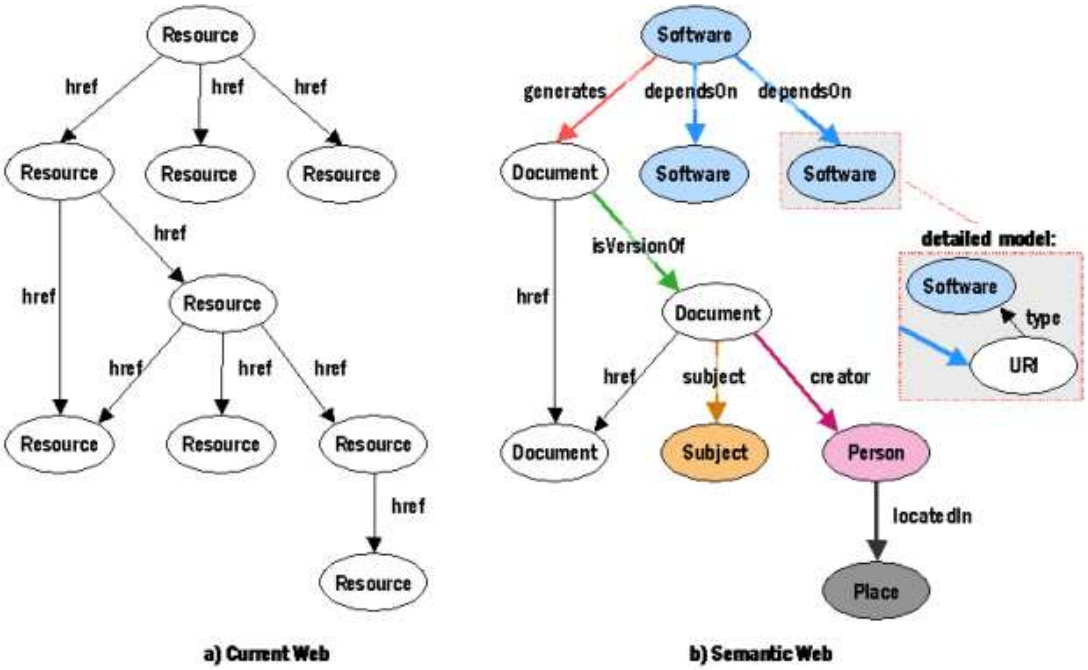
اما وب کنونی که امروزه در اختیار داریم و اصطلاحا به عنوان وب نحوی[4] شناخته مـی شـود، بـا ایـده اولیـه، تفاوت هایی دارد. اگر چـه در وب موجـود هـم همانطوریکـه در شـکل زیـر دیـده مـی شـود منـابع توسـط HyperLink ها به یکدیگر مرتبط می شوند ولی به دلیل استفاده گسترده از زبـان طبیعـی، عمـلا پـردازش‌ماشینی اطلاعات تحت الشعاع قرار گرفته است. در واقع، در وب فعلی تنها خود را محدود به یـک رابطـه ی خاص (HyperLink) بین منابع کرده ایم.

وب کنونی شامل منابع و لینک‌هاست. منابع اسناد وبی هستند برای مصرف انسان و به طور کلی شامل متا داده ای برای تشریح اینکه آن‌ها برای چه کاربردی هستند و روابط آن‌ها با سایر مستندات وب چیست ندارند. در حالی که یک انسان آگاه به سادگی می‌تواند درک کند که یک منبع خاص یک صورت حساب و یا یک رمان و یا یک مقاله تحقیقاتی است، اما این اطلاعات غالبا برای یک ماشین غیر قابل دسترس هستند. به طور مشابه یک کاربر با خواندن متن اطراف لینک‌ها می‌تواند حدس بزند که یک منبع چه نوع روابطی دارد، اما این حدس‌ها برای ماشین مشکل است[۴] .

همان گونه که در شکل شماره ۲ مشاهده می‌کنید، با این وجود که وب معنایی نیز شامل منابع و لینک‌ها است. هم چنین اکنون منابع و لینک‌ها می‌توانند نوع‌هایی داشته باشند که مفاهیمی را تعریف می‌کنند که اطلاعات بیشتری به ماشین‌ها می‌دهند. برای مثال بعضی لینک‌ها می‌گویند که

[3] Metadata
[4] Syntactical Web

یک منبع نسخه ای از منبع دیگری است یا توسط منبعی نوشته شده‌اند که یک شخص را توصیف می‌کند.




شکل ۲. وب کنونی در مقایسه با وب معنایی[۴]

بنابراین در وب موجود ما از کامپیوترها فقط به عنوان ابزاری جهت جمع آوری اطلاعات و نمـایش آن‌ها در قالب صفحات وب استفاده کرده و وظیفه تفسیر آنها را بر عهده انسان ها قرار می دهیم. با توجـه بـه گسترش روز افزون اطلاعات در وب، نیاز به استفاده بهتر از ماشین ها برای تفسیر و انجام اعمال بـر اسـاس اطلاعات روزانه، بیشتر می شود. سوال اساسی این است که چگونه می توان یـک وب از داده هـایی بـسازیم که قابل پردازش توسط ماشین باشد؟

پاسخ به سوال فوق را باید در روش نگرش خود به داده ها جستجو کنیم. تاریخ نشان می دهد کـه در ابتدا، هوشمندی، تماما در درون برنامه ها قرار داشت و (از نظر ساختاری) برنامه هـا کـاملا وابـسته بـه داده های خود بودند. در این نسل از برنامه ها، داده ها نقش ثانویه را در پردازش اطلاعات، بر عهـده داشـتند. بـا ظهور مدل شی گرایی در زبان های برنامه نویسی، داده ها استقلال بیشتری یافتند. ولـی هنـوز سـازندگان برنامه ها، بدلایل رقابتی و بازار، سعی می کردند تا همچنان داده های داخلی را وابسته به برنامه های خـود نگه دارند. در وب، به لطف وجود XML و ظهور وب

معنایی، داده ها نقش اول را ایفا مـی کننـد. از ایـن رو تلاش های زیادی صورت می گیرد تا هوشمندی را از درون برنامه ها به سمت داده ها منتقل نمایند. به این ترتیب می توان گفت که هدف اصلی در وب معنایی، ایجاد بستری جهت قابـل درک کـردن داده هـا بـرای ماشین (کامپیوتر) ها و ایجاد داده های هوشمندتر[۵] می باشد [۵].

## ۱.۲. طرح مساله

موتور جستجو یا جویشگر در فرهنگ رایانه به طور عمومی به برنامه ای گفته می‌شود که کلمات کلیدی را در یک سند یا بانک اطلاعاتی جستجو می‌کند، در اینترنت به برنامه ای گفته می‌شود که کلمات کلیدی موجود در صفحات وب جهانی، گروهای خبری، آرشیوها و ... را جستجو می‌کند. موتور های جستجوی مبتنی بر کلمات کلیدی، مهم‌ترین ابزار در پشتیبانی از بازیابی اطلاعات هستند. با توجه به انبوه اطلاعاتی که امروزه با آن مواجه هستیم، وجود این گونه ابزارها ضروری می‌باشد و بدون موتور های جستجو وب توفیقی ندارد. اما از طرفی دیگر، وب موجود، دارای یک مشکل اساسی می باشد. می توان گفت که چالش اساسی در ارتباط با وب آن است که منابع موجود تنها برای انسان ها قبل فهم و استفاده می باشد. بنابراین ماشین هـا در جستجو، اشتراک و یکپارچه سازی انبوهی از اطلاعات، محـدودیت هـای زیـادی دارنـد. در حـال حاضـر ماشین ها جهت شناخت محتوی و اطلاعات موجود در وب، نیاز به دانش انسانی دارند و بدون آن تنهـا مـی توانند کورکورانه، وجود یا عدم وجود یک کلمه کلیدی را مشخص نمایند.

یک دسته بندی از این مشکلات و محدودیت‌های اساسی برحسب فراخوانی[۶]، دقت و محتوا وجود دارند که بعضی از آن‌ها عبارتند از:

- فراخوانی کم[۷] یا هیچ: گاهی اوقات اتفاق می‌افتد که هیچ جوابی در پاسخ به درخواست وجود ندارد، و یا صفحات مرتبط و مهمی در فراخوانی‌ها یافت نمی‌شوند. اگر چه این حالت عموما کمتر رخ می‌دهد ولی غیر ممکن نیست.

---

[۵] منظور از هوشمندی داده ها، آن است که داده ها به تنهایی و بدون نیاز به نرم افزار مربوطه قابل استفاده بوده و بتوان دانش لازم را از اطلاعات موجود استنتاج نمود

[۶] یک فراخوانی، لینکی به صفحه ای از وب در پاسخ به درخواست کاربر از موتور جستجو است. این فراخوانی می تواند شامل صفحه ای مرتبط، کم مرتبط و یا نا مرتبط به موضوع درخواست باشد.

[7] Low Recall

- فراخوانی زیاد[8] و دقت کم: تعداد نتایج حاصل از جستجو بسیار زیاد است و به علاوه این نتایج شامل صفحات کم مرتبط و نامرتبط هم می‌شود، در نتیجه یافتن اطلاعات مورد نظر از لیست نتایج جستجو بسیار مشکل خواهد بود.
- نتایج به لغات وابستگی زیادی دارند: موارد زیادی اتفاق می‌افتد که کلمه کلیدی مورد جستجوی شما معانی متفاوتی داشته باشد، که موتور های جستجوی کنونی نمی‌توانند تفاوتی بین آن‌ها قائل شوند مثلا کلمه «شیر» که در کاربرد های متفاوت معانی متفاوتی دارد، گاهی اوقات هم تشخیص اختلاف معنا برای موتور جستجو غیر ممکن است، به عنوان مثال تشخیص تفاوت معنای دو جمله متضاد «من دانشجوی رشته کامپیوتر هستم» و «شما ممکن است فکر کنید **من دانشجوی رشته کامپیوتر هستم** ولی این طور نیست» برای این جستجوگرها ممکن نیست. به طوری که کامپیوتر هنگام بازیابی، اختلاف قابل ملاحظه ای بین این دو متن قائل نمی‌شود، در حالی که کاربر با وجود مشابهت زیاد این دو عبارت به خوبی قادر به تشخیص تضاد آن‌هاست [۶].

این‌ها تنها چند مورد از مشکلات موجود در موتور های جستجوی فعلی می‌باشند. این گونه مشکلات در سایر ابزار های مورد استفاده در وب هم نمود می‌کند. استفاده از وب معنایی راهکار متداول برای حل مشکلات اشاره شده می‌باشد.

آقای تیم برنرزلی وب معنایی را به عنوان وب داده‌ها[9] معرفی کرده است و به این نکته اشاره کرده است که وب معنایی یا نسخه سوم وب[10] در مورد اتصال دادها است. وب معنایی وابسته به ارتباط دادن معانی به محتوای داده می‌باشد که این مهم از طریق آنتولوژی ها[11] -فرهنگ‌های لغات رسمی برای تعریف کلمات از طریق اعلان مفاهیم و ارتباطات بین آنها به شیوه ای غیر مبهم می‌باشند- قابل انجام است. پس مهم‌ترین نکته در عملیاتی نمودن وب معنایی وجود داده‌هایی با ساختاردهی مورد انتظار است و استخراج و تعریف آنتولوژی ها برای این ساختارها می‌باشد [۵].

[8] High Recall
[9] Web of Data
[10] Web 3.0
[11] Ontology

مساله مطرح در این پایان نامه ارائه یک روش کلی برای عملیاتی نمودن پشتیبانی از جستجوی معنایی در یک سازمان می‌باشد. در راستای این هدف چالشهای زیر مطرح است:

۱. ارائه روشی برای ساختار دهی داده‌ها و استخراج آنتولوژی از داده های موجود.
۲. طراحی موتور جستجوی مبتنی بر آنتولوژی.
۳. اجرای روش مورد نظر بر روی یک مورد مطالعه.

## ۱.۳. اهداف تحقیق

نمایش نتایج جستجو به این موضوع که از ابتدا چه اطلاعاتی باید نمایش داده شوند وابسته است. با این فرض که نسل جاری موتور های جستجو درک بسیار محدودی از پرس و جو[۱۲] انجام شده توسط کاربر و درک منظور واقعی کاربر دارند، شانس داشتن نتایج ایده آل کمتر می‌شود.

امروزه بسیاری از صفحات وب اطلاعاتی را که نمایش می‌دهند، از پایگاه داده های خود استخراج می‌کنند که عموما به آن‌ها صفحات وب پویا[۱۳] گفته می‌شود. در مورد برنامه های کاربردی نیز به همین منوال است و اکثر آن‌ها اطلاعات خود را در پایگاه داده ذخیره می‌کنند. استفاده از پایگاه داده باعث ساختار دهی مناسب اطلاعات می‌شود.

حال فرض کنید در موتور جستجو کاربر در هنگام ورود پرس و جوی مورد نظر خود، منظور خود و اینکه دقیقا چه می‌خواهد، را در موتور جستجو وارد کند، مثلا اینکه او به دنبال یک **شهر** است که در کشور **ایران** قرار دارد که **جمعیت** آن ۲ میلیون نفر است. همچنین فرض کنید اطلاعاتی که موتور جستجو روی آن‌ها جستجو انجام می‌دهد به شیوه ای ساختار دهی شده باشند و حاوی فرا داده‌هایی[۱۴] باشند تا مفاهیم، ویژگی‌ها و روابط بین آن‌ها (سه جز اصلی آنتولوژی) با هم مشخص شده باشد، حال با در اختیار داشتن چنین زیر ساختی، موتور جستجو می‌تواند کاربر را به همان نتایجی برساند که دقیقا به دنبال آن‌ها بوده است. ایده اصلی این تحقیق بر اساس فرضیات بالا استوار است.

[12] Query
[13] Dynamic Web Pages
[14] Meta Data

اما سوالاتی که می‌توان مطرح کرد عبارتند از:

- آیا می‌توان یک متد کلی برای سازمانهای مختلف ارائه داد که بتوانند جستجوی معنایی در پایگاه داده های درون سازمانی خود داشته باشند؟
- یک موتور جستجو چه میزان می‌تواند به صورت قابل استفاده مجدد طراحی شود که برای هر دامنه ای از مساله مورد استفاده قرار گیرد؟
- آیا در متد ارئه شده تنوع تکنولوژی هم مطرح است؟ به عنوان مثال آیا فقط بر روی پایگاه داده رابطه ای[۱۵] به عنوان نمونه SQL، قابل پیاده سازی است و یا بر روی هر نوع ساختار اطلاعاتی از جمله فایل، XML و سایر ساختارها قابل پیاده سازی می‌باشد.
- آیا راهکار ارائه شده هم بر روی برنامه های کاربردی و هم وب سایتهای پویا قابل استفاده خواهد بود؟
- و به طور کلی پیش نیازها، محدودیت‌ها و حوزه های استفاده کدامند؟

## ۱.۴. روش انجام تحقیق و دستاوردها:

### ۱.۴.۱. مراحل تحقیق:

این تحقیق به گونه ای برنامه ریزی شده است که به غیر از پرداختن به مباحث نظری وارائه الگو، یک نمونه کاربردی واقعی نیز پیاده سازی شود. پس به طور کلی این تحقیق می‌تواند شامل دو فاز باشد که یکی شامل مباحث تئوریک و بررسی تکنولوژی‌ها و ابزارها و ارائه الگو می‌باشد و دیگری طراحی و پیاده سازی نمونه عملیاتی می‌باشد. البته واضح است که لزوما تمامی مراحل فاز دوم بعد از اتمام فاز اول نمی‌باشد.

مراحلی که در هر یک از فازها می‌بایست انجام شوند بدین قرارند:

- **فاز نخست:**
  - ❖ مرحله ۱- بررسی مفاهیم وب معنایی، آنتولوژی، زبان توصیف منابع RDF[۱۶]، زبان آنتولوژی OWL[۱۷] ، زبان بازجست SPARQL[۱۸]

[15] RDB(relational data base)
[16] Resource Description Framework
[17] Web Ontology Language --http://en.wikipedia.org/wiki/Web_Ontology_Language