بسم الله الرحمن الرحیم

دانشگاه شیخ بهایی

دانشکده مهندسی

پایان‌نامه کارشناسی ارشد رشته مهندسی کامپیوتر- نرم افزار

تولید خودکار آنتولوژی برای کدهای برنامه‌نویسی تحت وب

پژوهشگر

آربی باغومیان

استاد راهنما

دکتر ناصر قاسم آقایی

شهریور ۱۳۹۲

باسمه تعالی

دانشگـــــــــاه شیـــــــــخ بهائی

دانشکده مهندسی

گروه مهندسی کامپیوتر و فناوری اطلاعات

پایان‌نامه کارشناسی ارشد رشته مهندسی کامپیوتر – نرم‌افزار آقای آربی باغومیان

تحت عنوان

تولید خودکار آنتولوژی برای کدهای برنامه‌نویسی تحت وب

در تاریخ **۱۳۹۲/۰۷/۲۵** توسط هیأت داوران زیر بررسی و با درجه به تصویب نهائی رسید.

۱- استاد راهنمای پایان‌نامه **دکتر ناصر قاسم آقایی** با مرتبه علمی **استاد** امضاء

۲- استاد داور داخل گروه **دکتر احمد براآنی** با مرتبه علمی **دانشیار** امضاء

۳- استاد داور خارج گروه **دکتر محمد حسین ندیمی** با مرتبه علمی **استادیار** امضاء

دکتر سید محمدحسن فیض

مدیر تحصیلات تکمیلی

سپاسگزاری

این پایان‌نامه فرصتی است که از پدر و مادرم و همچنین کمک‌های بی‌شائبه خواهرم تشکر و قدردانی کنم، همچنین از اساتید گرانقدر جناب پروفسور ناصر قاسم آقائی که علیرغم مشغله کاری بسیار اینجانب را در این مسیر یاری رساندند و همچنین از آقای دکتر احمد براآنی بابت راهنمایی‌های سازنده ایشان برای اتمام این پایان‌نامه.

در آخر از یار و همدم همیشگی خود بابت همراهی در تمام مراحل زندگی خود، کمال قدردانی و تشکر را دارم.

بی‌شک بدون همیاری و بردباری پدر و مادر عزیزم هیچگاه به این مرحله از زندگی خود نمی‌رسیدم، امیدوارم در ادامه این راه حضور گرم و دلگرم‌کننده آن‌ها را در کنار خود احساس کنم .

چکیده

در سال ۲۰۰۶ تیم برنرز-لی ایده وب آینده را به نام وبی از داده ها مرتبط مطرح کرد. این ایده تکمیل شده ایده خود او در سال ۲۰۰۱ با نام وب معنایی بود که عبارت بود از : وبی تشکیل شده از داده‌هایی که توسط ماشین‌ها بصورت مستقیم یا غیر مستقیم ، قابل پردازش می‌باشد. با معرفی این ایده انگیزه برای منطبق کردن وب‌سایت‌ها، برنامه‌های تحت وب و هر آنچه در این فضا می‌باشد ایجاد کرد. ایده ارائه شده به گونه‌ای بود که هر حوزه از دانش باید به طور جداگانه خود را با این محیط منطبق کند. برای انطباق با این محیط باید بر روی برنامه‌های تحت وب زبانی را ایجاد کرد که به واسطه آن ماشین‌ها بتوانند این برنامه‌ها را درک کنند، نام این زبان آنتولوژی می‌باشد.

با مطرح شدن این ایده تمام حوزه‌های دانش شروع به افزودن آنتولوژی به برنامه‌های خود شدند. عیب موجود در این می‌باشد که افزودن دستی آنتولوژی بر روی برنامه‌های بزرگ تحت وب کاری بسیار زمانبر می‌باشد و احتمال خطا در آن بسیار زیاد می‌باشد. به همین دلیل هر حوزه از دانش شروع به اعمال مکانیزمی جهت تولید خودکار آنتولوژی پرداختند.

ایده ارائه شده در این پروژه در مورد اعمال مکانیزمی جهت تولید خودکار آنتولوژی برای کدهای برنامه‌نویسی تحت وب می‌باشد. با اعمال این ایده ماشین‌هایی همچون موتورهای جستجو می‌توانند کدهای برنامه‌نویسی را همانند انسان درک کنند. نتیجه‌ای که این ایده دارد در این می‌باشد که کاربران می‌توانند پروژه‌های مورد نظر خود را در فضای مجازی بسیار بزرگ وب با دقت بهتری جستجو کنند و مجموعه‌ای از پروژه‌هایی را که نزدیک به خواسته خود می‌باشد، پیدا کنند.

پس از تولید آنتولوژی جهت بررسی صحت آنتولوژی تولیدی از لحاظ رعایت قوانین آنتولوژی آن را توسط مرجع رسمی بررسی صحت آنتولوژی که همان کنسرسیوم دبلیو-۳ می‌باشد ارزیابی می‌شود. در ارزیابی انجام شده نشان داده شد که آنتولوژی تولیدی از لحاظ رعایت قوانین نگارشی مربوط به آنتولوژی معتبر بوده است.

کلیدواژه :

آنتولوژی، تولید خودکار آنتولوژی، داده‌های پیوندی، وب‌معنایی، پایگاه‌داده سه-تایی

فهرست مطالب

فصل اول: کلیات ۱

۱-۱بیان مساله پژوهشی ۲

۱-۲رهیافت پیشنهادی ۶

۱-۳ اهداف و کاربرد پژوهش ۷

۱-۴چارچوب پایان‌نامه ۷

فصل دوم: ادبیات تحقیق ۹

۲-۱ مقدمه ۱۰

۲-۲ انگیزش و توضیح مساله ۱۰

۲-۳ وب دو چیست ؟ ۱۱

۲-۴ وب دو بعنوان یک پلتفرم ۱۲

۲-۵ تعاریف پایه‌ای: وب دو ، وب یک و وب سه ۱۳

۲-۵-۱ تعریف جامع ۱۳

۲-۵-۲ سبدهای خرید از نوع وب یک ۱۳

۲-۵-۳ دیدگاههای مختلف درباره وب دو ۱۵

۲-۵-۴ وب سه ۱۶

۲-۵-۵ تغییرات در وب دو و وب سه ۱۶

۲-۵-۶ نت‌اسکیپ در مقابل گوگل ۱۸

۲-۶ وب داده ۲۰

۲-۷ قوانین آدرس‌دهی اینترنتی ۲۱
۲-۷-۱ یو-آر-آی‌های ۳۰۳ ..... ۲۲
۲-۷-۲ هش یو-آر-آی ۲۳
۲-۸ جمع‌بندی ۲۴
فصل سوم: پیشینه تحقیق ۲۵
۳-۱مقدمه ۲۶
۳-۲ تولید خودکار آنتولوژی از وب عمیق : ۲۷
۳-۲-۱ مرور کلی بر ایده ارائه شده ۲۸
۳-۲-۲ استخراج ویژگی‌ها از وب عمیق ۲۸
۳-۲-۳ ایجاد آنتولوژی از وب عمیق ۲۹
۳-۳ استخراج خودکار دانش از مستندات وب : ۳۰
۳-۴ تولید خودکار آنتولوژی با استفاده از کلیدواژگان جستجو: ۳۲
۳-۴-۱ تولید خودکار آنتولوژی با استفاده از واژگان ورودی ۳۳
۳-۴-۱-۱ آنالیز عبارات ورودی ۳۳
۳-۴-۱-۲ آنالیز معنای واژگان ۳۳
۳-۴-۱-۳ ایجاد ساختار سلسله مراتبی با استفاده از اف-سی-ای: ۳۳
۳-۵ روش سبک برای تولید و تغییر آنتولوژی همزمان با تغییر منابع پایگاه‌داده رابطه‌ای : ۳۴
۳-۵-۱ تولید آنتولوژی از پایگاه‌داده ۳۵
۳-۵-۲ مدل گراف مفهومی از قوانین نگاشت آنتولوژی ۳۶
۳-۵-۳ تغییر شناخت و همزمانی بین پایگاه‌داده و آنتولوژی ۳۷
۳-۵-۴ مجموعه تغییرات ۳۸
۳-۶ تولید خودکار آنتولوژی برای مقالات عربی ویکیپدیا : ۴۱
۳-۶-۱ پیش‌زمینه ۴۱

۳-۶-۲ شرح سیستم ۴۲
۳-۷ جمع‌بندی ۴۴
فصل چهارم: مدل پیشنهادی تولید خودکار آنتولوژی برای کدهای برنامه‌نویسی تحت وب..... ۴۵
۴-۱ مقدمه ۴۶
۴-۲ فرآیند تولید آنتولوژی..... ۴۶
۴-۳ معماری سیستم ۴۸
۴-۴ کاوش ۴۸
۴-۵ استخراج اطلاعات ۴۹
۴-۶ تولید آنتولوژی ۵۱
۴-۷ جمع‌بندی..... ۵۱
فصل پنجم: پیاده‌سازی مدل پیشنهادی..... ۵۳
۵-۱ مقدمه ۵۴
۵-۲ کاوش..... ۵۶
۵-۳ استخراج اطلاعات..... ۵۹
۵-۴ تولید آنتولوژی ۶۸
۵-۴-۱ آنتولوژی..... ۶۸
۵-۴-۲ پایگاه‌داده سه-تایی آرک-۲..... ۷۰
۵-۴-۳ الگوریتم تولید آنتولوژی..... ۷۴
۵-۴-۳-۱ تعاریف اولیه فایل آنتولوژی :..... ۷۴
۵-۴-۳-۲ آنتولوژی توضیحات پروژه :..... ۷۸
۵-۴-۳-۳ آنتولوژی کدهای پروژه :..... ۷۹
۵-۴-۳-۴ پرس‌وجو از پایگاه‌داده سه-تایی آرک-۲ :..... ۸۱

۵-۴-۳-۴-۱ اتصال به پایگاه‌داده : ۸۱
۵-۴-۳-۴-۲ ثبت عبارات سه-تایی : ۸۱
۵-۴-۳-۴-۳ بازیابی از پایگاه‌داده سه-تایی : ۸۲
۵-۴-۳-۵ الگوریتم ارزیابی شباهت : ۸۳
۵-۵ جمع‌بندی ۸۸
فصل ششم: ارزیابی ۸۹
۶-۱ مقدمه ۹۰
۶-۲ ارزیابی صحت آنتولوژی تولید شده ۹۰
۶-۳ ارزیابی بر اساس پروژه‌های واقعی ۹۱
۶-۴ جمع بندی ۹۷
فصل هفتم: نتیجه و کارهای آینده ۹۸
۷-۱ نتیجه‌گیری ۹۹
۷-۲ کارهای آینده ۹۹
منابع و مآخذ ۱۰۱

فهرست شکل‌ها

شکل (۱-۱): نمونه آنتولوژی ۴

شکل (۱-۲): محیط انتخابی برای پروژه ۸

شکل (۲-۱): تغییرات ایجاد شده در برنامه‌های وب-۲ نسبت به وب-۱ ..... ۱۲

شکل (۲-۲): نقشه الگوی وب دو ۱۴

شکل (۲-۳): هدر به فرمت آر-دی-اف/ایکس-ام-ال ۲۲

شکل (۲-۴): هدر به فرمت متن/ اچ-تی-ام-ال ۲۲

شکل (۲-۵): قالب هش یو-آر-آی ۲۳

شکل (۲-۶): یو-آر-آی با استراتژی هش ۲۳

شکل (۳-۱): پردازش فرآیندها. ۲۸

شکل (۳-۲): مثالی از تولید یک قطعه شمای بر اساس نزدیکترین والد مشترک "شخص". ۳۰

شکل (۳-۳): مدلی از بیوگرافی تولید شده توسط سیستم. ۳۲

شکل (۳-۴): ساختار سلسله مراتبی تولید خودکار آنتولوژی با استفاده از واژگان ورودی ۳۴

شکل (۳-۵): ارتباط دو نود به صورت مستقیم. ۳۵

شکل (۳-۶): فرایند ساخت مجموعه تغییرات. ۴۰

شکل(۳-۷): جدول اطلاعات موجود در مقالات ویکیپدیا. ۴۱

شکل (۳-۸): ارتباط چهار کلاس سیستم. ۴۲

شکل (۳-۹): ساختمان سیستم تولید آنتولوژی از مقالات عربی ویکیپدیا ۴۳

شکل (۴-۱): رابط کاربری جهت ورود اطلاعات پروژه ۴۶

شکل (۴-۲): نحوه دریافت فایل آنتولوژی تولیدی از سیستم ۴۷

شکل (۴-۳): معماری مدل پیشنهادی ۴۸

شکل (۴-۴) فلوچارت کاوش ۴۹

شکل (۴-۵): فلوچارت استخراج اطلاعات ۵۰

شکل (۴-۶): فلوچارت تولید آنتولوژی ۵۲

شکل (۵-۱): زبان برنامه‌نویسی استفاده شده در وب‌سایت‌ها ۵۵

شکل (۵-۲): سیستم‌عامل استفاده شده در وب سایت‌ها ۵۵
شکل (۵-۳): وب سرور استفاده شده در وب‌سایت‌ها..... ۵۶
شکل (۵-۴): ساختار مدل سه لایه‌ای ام-وی-سی. ۵۷
شکل (۵-۵): شبه کد کاوش..... ۵۹
شکل (۵-۶): ساختار یو-آر-آی ۶۰
شکل (۵-۷): مقایسه تعداد یو-آر-آی در مقابل تعداد یو-آر-ال ۶۱
شکل (۵-۸): نمونه فایل ایکس-ام-ال..... ۶۴
شکل (۵-۹): اطلاعات استخراج شده به فرمت متن ۶۶
شکل (۵-۱۰): نحوه دریافت اطلاعات از کلیدواژه‌های اچ-تی-ام-ال..... ۶۷
شکل (۵-۱۱): شبه کد استخراج اطلاعات ۶۷
شکل (۵-۱۲): دستورات اُ-دبلیو-ال ۶۹
شکل (۵-۱۳): جدول حاوی مقادیر موضوع..... ۷۱
شکل (۵-۱۴): جدول حاوی مقادیر ویژگی..... ۷۲
شکل (۵-۱۵): جدول حاوی مقادیر شی ۷۲
شکل (۵-۱۶): جدول حاوی سه-تایی موضوع، ویژگی و شی ۷۳
شکل (۵-۱۷): نمونه تعریف منبع ۷۵
شکل (۵-۱۸): تعاریف اولیه فایل آنتولوژی..... ۷۵
شکل (۵-۱۹): تعریف کلاس در آنتولوژی..... ۷۶
شکل (۵-۲۰): معرفی ویژگی وجود یک کلاس در یک فایل ۷۶
شکل (۵-۲۱): معرفی ویژگی نام متد ۷۷
شکل (۵-۲۲): معرفی ویژگی تعداد متد ۷۷
شکل (۵-۲۳): معرفی ویژگی تعداد کلاس ۷۸
شکل (۵-۲۴): دستور جستجو در پایگاه‌داده سه-تایی ۷۸
شکل (۵-۲۵): کد آنتولوژی مربوط به توضیحات پروژه..... ۷۹
شکل (۵-۲۶): ساختار کلی آنتولوژی کدها ۸۰

شکل (۵-۲۷): مثالی از نودهای خالی ۸۱
شکل (۵-۲۸): عبارات سه-تایی مربوط به کلاس ۸۱
شکل (۵-۲۹): عبارت سه-تایی مربوط به پروژه ۸۲
شکل (۵-۳۰): دستور مربوط به بازیابی از پایگاه‌داده سه-تایی مربوط به پروژه ۸۳
شکل (۵-۳۱): دستور مربوط به بازیابی از پایگاه‌داده سه-تایی مربوط به کلاس ۸۳
شکل (۵-۳۲): ساختار سه-تایی نام کلاس ۸۴
شکل (۵-۳۳): ساختار سه-تایی توضیحات کلاس ۸۵
شکل(۵-۳۴): ساختار سه-تایی نام توسعه دهندگان کلاس ۸۵
شکل (۵-۳۵): نمونه آنتولوژی تولید شده ۸۶
شکل (۵-۳۶): شبه کد تولید آنتولوژی ۸۷
شکل (۶-۱): فرآیند وارد کردن اطلاعات جهت اعتبارسنجی آنتولوژی تولید شده ۹۰
شکل (۶-۲): پیغام صحت مبنی بر معتبر بودن آنتولوژی تولید شده ۹۱
شکل (۶-۳): تعاریف اولیه فایل آنتولوژی تولید شده مربوط به پروژه مرورگر الستیک ۹۲
شکل (۶-۴):تعاریف اولیه کلاس‌ها و ویژگی‌ها مربوط به پروژه مرورگر الستیک ۹۲
شکل (۶-۵): ارتباط معنایی تولید شده مربوط به پروژه مرورگر الستیک ۹۳
شکل (۶-۶): صحت عدم وجود کلاس متشابه در پروژه هسته الستیک ۹۴
شکل (۶-۷): تعاریف اولیه فایل آنتولوژی تولید شده مربوط به پروژه هسته الستیک ۹۴
شکل (۶-۸): تعاریف اولیه کلاس‌ها و ویژگی‌ها مربوط به پروژه هسته الستیک ۹۵
شکل (۶-۹) ارتباط معنایی تولید شده مربوط به پروژه هسته الستیک ۹۵
شکل (۶-۱۰) صحت عدم وجود کلاس متشابه در پروژه مرورگر الستیک ۹۶

فهرست جدول‌ها

جدول (۳-۱): جدول گراف مفهومی از قوانین نگاشتی. ۳۶
جدول (۳-۲): قوانین نگاشت آنتولوژی. ۳۷
جدول (۳-۳): قوانین شناخت تغییرات. ۳۹

فصل اول:

کلیات

۱-۱ بیان مساله پژوهشی

انسان‌ها جهت برقراری ارتباط با یکدیگر از یک زبان مشترک استفاده می‌کنند، این کار باعث می‌شود همدیگر را بهتر درک کنند و واکنش‌های مناسبتری را نسبت به یکدیگر انجام دهند. یکی از دلایلی که ماشین‌ها عکس‌العمل مناسبی نسبت به خواسته انسان‌ها انجام نمی‌دهند عدم وجود زبان مشترک بین ماشین و انسان می‌باشد. به طور مثال موتورهای جستجوگر وب اطلاعاتی را بر می‌گرداند که بیشتر آن‌ها ارتباطی با خواسته کاربر نداشته و تنها به صرف داشتن ارتباط جزئی آن را به عنوان خواسته کاربر بر می‌گردانند. به همین منظور تیم برنرز-لی[1] خالق وب مفهومی را به نام داده‌های پیوندی مطرح کرد که به واسطه آن ماشین‌ها می‌توانند همانند انسان مفاهیم را درک کنند و واکنشی مناسب نسبت به آن انجام دهند. زبان مشترک بین انسان و ماشین آنتولوژی می‌باشد. در دنیای امروزی به دلیل ماهیت جدا بودن حوزه‌های دانش از یکدیگر برای هر حوزه از دانش آنتولوژی مختص به آن دانش را ایجاد می‌کنند -مانند رشته کامپیوتر که اصطلاحات مختص به خود را نسبت به سایر دانش‌ها دارد.

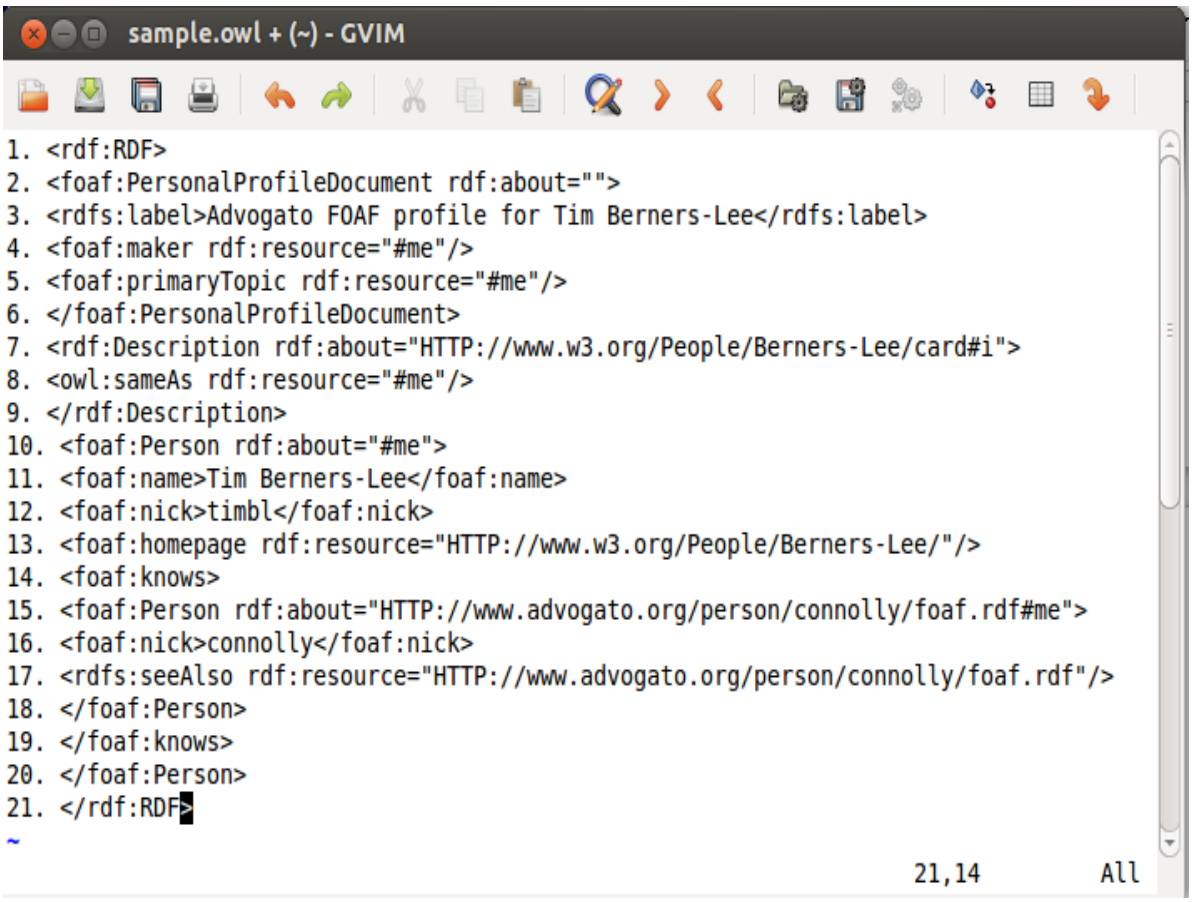
در سالهای اخیر تحقیق برای پیشبرد محیط وب به طور قابل توجهی گسترش پیدا کرده است. یکی از کاستی‌های موجود بر اساس تقاضای کاربران می‌باشد، به طور مثال می‌توان به پیدا کردن اطلاعات و یا دانش مورد نیاز برای کاربران اشاره کرد.

به این معنا که آسان‌ترین راه برای اینکه کاربران به اطلاعات مورد نیاز خود دست پیدا کنند استفاده از موتورهای جستجو مانند گوگل، بینگ، یاهو و غیره می‌باشد، اما مشکل اصلی آن‌ها حجم زیاد اطلاعات برگردانده شده بر اساس کلید واژه مورد نظر کاربر می‌باشد که اکثر این نتایج، خواسته کاربر نمی‌باشد و تنها به صرف اینکه برخی از کلید واژه‌ها در اطلاعات جستجو شده یافت شده به عنوان نتایج مورد نیاز برگردانده می‌شود و تنها یک نوع سردرگمی برای کاربر ایجاد می‌کند. این نوع جستجو در تمام زمینه‌ها وجود دارد (موسیقی، فیلم، مستندات و حتی کدهای برنامه‌نویسی). در زمینه برنامه‌نویسی کدها، همیشه از آن‌ها جهت بکارگیری مجدد استفاده می‌شدند اما یکی از اولین پیشنهاداتی که بطور خاص از کدها جهت ایجاد قابلیت استفاده مجدد ارائه شد توسط پروفسور مک-روی[2] بود که در سال ۱۹۶۸ نرم‌افزارها را بر اساس قابلیت استفاده مجدد از کامپوننت‌ها ارائه کرد. یکی از اهداف اصلی پیدایش زبان‌های مبتنی بر شی‌گرایی قابلیت استفاده مجدد از آن‌ها بود که با پیدایش زبان‌های مختلف در محیط‌های گوناگون (اینترنت، شبکه، هوش‌مصنوعی، سیستم‌عامل و غیره) استفاده از این قابلیت به طور چشم‌گیری افزایش یافت به طوری که با پیدایش چهارچوب‌های مختلف برای زبان‌های برنامه‌نویسی دیگر نیازی به نوشتن کدهای پایه نبود و تنها تمرکز بر روی برآورده کردن درخواست‌های مشتری معطوف می‌شد.

حال با گسترش استفاده از قابلیت استفاده مجدد و همین طور استفاده از موتورهای جستجو برای پیدا کردن اطلاعات

[1] Tim berners-Lee
[2] Malcolm Douglas McIlroy

مورد نیاز به نظر می‌رسید مشکل پیدا کردن کدها و یا پروژه‌های موردنیاز حل شده باشد اما به علت مشکلی که در موتورهای جستجو (که در بالا به آن ذکر شد) وجود دارد، باعث می‌شود که بازدهی و دقت نتایج جستجو شده کاهش پیدا کند و کاربر، پروژه‌ها و یا کدهای مناسبی را پیدا نکند در حالی که اگر کمی کلید واژه خود را تغییر دهد ممکن است نتایجی متفاوت و نزدیکتر به خواسته خود پیدا کند که همین کار باعث بالا رفتن زمان جستجو و حتی نتیجه نگرفتن شود. برای حل این مشکل تیم برنرز–لی خالق وب، اصطلاحی به‌نام وب معنایی را تعریف کرد که به معنای استفاده مستقیم و یا غیر مستقیم ماشین از داده‌های وب می‌باشد [۱، ۲]. هدف از ایجاد وب‌معنایی درک ماشین‌ها از اطلاعات و ارتباط کلمات با هم بود تا ماشین‌ها بتوانند مفهوم کلمات را مانند انسان درک کنند. این کاربرد در موتور جستجو باعث می‌شود که جستجوگر وب تنها بر اساس واژه مورد نظر به جستجو نپردازد بلکه بتواند معنای عبارت جستجو را درک کند و به دنبال اطلاعاتی در همین رابطه برود نه به صرف هماهنگی در کلمات موجود در عبارت جستجو. پیدایش این تکنولوژی باعث شد تا دقت و بازدهی جستجو چندین برابر شود گرچه هنوز به طور عام از این تکنولوژی استفاده نمی‌شود. یکی از ویژگی‌های اصلی وب‌معنایی ایجاد قابلیت اشتراک‌گذاری و استفاده مجدد می‌باشد. همین ویژگی بسیاری از محققان و کاربران را به استفاده و تحقیق بر روی این تکنولوژی جلب کرده است. افزودن این تکنولوژی بر روی وب کنونی باعث ارتقا آن از لحاظ قابلیت اشتراک‌گذاری و استفاده مجدد می‌شود به طوری که می‌توان کدها و پروژه‌ها را با بالاترین دقت و نزدیک به عبارت جستجویی مورد نظر خود بدست آورد، اما یک مشکل بر سر این راه وجود دارد و آن این است که زمانی وب‌معنایی کارایی دارد که مفاهیم صفحات و اطلاعات موجود در وب ابتدا استخراج شده باشد چراکه وب‌معنایی بر اساس داده‌های آنتولوژی می‌تواند چنین قابلیتی را ایجاد کند. در واقع آنتولوژی زبان واسط میان انسان و ماشین‌ها می‌باشد بدین معنا که ماشین‌ها جهت درک موضوع، اطلاعات موجود در آنتولوژی را می‌خوانند و بر اساس آن درکی از اطلاعات مرتبط با آنتولوژی کسب می‌کنند. به طور مثال اگر شخصی بخواهد اطلاعات شخصی خود را در سایت خود قرار دهد و اطلاعاتی از قبیل نام مستعار، افرادی را که در ارتباط با کار خود می‌باشند و همینطور اطلاعات دیگر از سایت‌های دیگر که مرتبط با موضوعات خود می‌باشد ارائه دهد و بخواهد این اطلاعات برای ماشین قابل فهم باشد به این مفهوم که ماشین‌ها هنگام بازدید از سایت فرد بتوانند این اطلاعات را کسب کنند. به طور مثال ماشین‌ها بتوانند افرادی که صاحب سایت آن‌ها را می‌شناسد مشخص کنند باید از آنتولوژی استفاده کند. به عنوان نمونه در شکل (۱–۱) یک آنتولوژی ارائه شده توسط خود تیم برنرز لی می‌باشد که به معرفی خود پرداخته است. همانطور که مشخص است از خط ۱ تا ۱۳ به معرفی خود پرداخته است که شامل نام خود، نام مستعار و آدرس صفحه اینترنتی شخصی خود می‌باشد و از خط ۱۴ به بعد فردی را که می‌شناسد معرفی کرده است. ماشین‌ها با خواندن یک فایل آنتولوژی همانند



شکل (۱-۱): نمونه آنتولوژی

شکل (۱-۱) می‌توانند در مورد اطلاعات قرار داده شده در صفحه اینترنتی درکی پیدا کنند این کار به ماشین‌ها کمک می‌کند که واکنش مناسبتری را نسبت به خواسته کاربران ارائه دهند.

از زمان ظهور پروژه وب‌معنایی بارها تفسیرهای متفاوتی از آن شده است، اما هدف اصلی آن، ایجاد یک وب جهانی قابل خواندن توسط ماشین‌ها می‌باشد، تغییر نکرد. به همین دلیل برنرز-لی وب‌معنایی را به معنای استفاده مستقیم و یا غیر مستقیم ماشین از داده‌های وب تعریف کرد. از همین رو، وب معنایی و یا داده‌های وب هدف اصلی این فرایند می‌باشد، اما داده‌های پیوندی راه را برای رسیدن به این هدف فراهم می‌کنند [۳]. داده‌های پیوندی بر اساس چهار قانون می‌باشند که خالق آن برنرز-لی آن را معرفی کرد.

این چهار قانون عبارتند از [۴] :

۱. برای نامگذاری و یا شناسایی اشیا، موجودات و غیره از یو-آر-آی استفاده کنید.

۲. از پروتکل اچ-تی-تی-پی جهت تبادلات اطلاعات استفاده کنید تا برای انسان و ماشین‌ها قابل شناسایی و رویت باشند.

۳. از استانداردهایی مانند آر-دی-اف و اسپارکیو-ال جهت ارائه اطلاعات کاربردی به کاربران استفاده کنید.

٤. بین لینک‌های مرتبط به هم، ارتباط برقرار کنید.

در واقع وب‌معنایی را به عنوان یک کل و داده‌های پیوندی را بخشی از آن می‌دانند به عبارت دیگر وب معنایی را "چه‌چیز" و داده‌های پیوندی را "چگونگی" رسیدن به این هدف تعریف کرده‌اند. بطوریکه تیم برنرز-لی در سمیناری که در سال ۲۰۰۸ ارائه داد اعلام کرد که داده‌های پیوندی، "هدف وب‌معنایی را بدرستی انجام داد" [۳] .

اما ممکن است سوال شود که در عمل داده‌های پیوندی چه کمکی به پیشبرد اهداف این پروژه می‌کند؟

داده‌های پیوندی با ایجاد شبکه‌ای گسترده از داده‌های به هم‌مرتبط از لحاظ محتوا باعث می‌شود جستجو تنها بر روی هدف مورد نظر معطوف شود و دیگر داده‌های غیرمرتبط که تنها به صرف یک شکل بودن کلمات در جستجو پدیدار می‌شوند از لیست محتویات جستجو شده حذف و یا به آخر این لیست از لحاظ اهمیت موضوع و مرتبط بودن با موضوع مورد نظر برود.

برای دستیابی به این هدف باید چهار قانون داده‌های پیوندی را که در بالا به آن ذکر شد اعمال کرد. برای این کار باید بر روی صفحات وب اطلاعاتی علاوه بر داده‌های موجود در آن اضافه شود که این اطلاعات چیزی جز آنتولوژی نمی‌باشد، در آنتولوژی موجودیت‌ها بر اساس یو-آر-آی شناسایی و نام‌گذاری می‌شوند و همچنین می‌توان برای درک بهتر عامل‌ها و یا همان ماشین‌ها، به سایر لینک‌های مرتبط اتصالی برقرار کرد ، برای ارتباط نیز از پروتکل اچ-تی-تی-پی استفاده می‌کند و زبان آن می تواند آر-دی-اف، اُ-دبلیو-ال و از این دست باشد [۵، ۶]. به عنوان مثال یک نمونه از آنتولوژی‌های ایجاد شده که مربوط به مشخصات تیم برنرز-لی می‌باشد در شکل (۱-۱) نشان داده شده است.

همانطور که مشخص است در شکل (۱-۱) واژگانی چون sameAs، seeAlso، foaf و غیره دیده می‌شود. واژه sameAs به ماشین‌ها می‌گوید که لینک مورد نظر از لحاظ محتوا همانند صفحه کنونی می‌باشد و یا seeAlso به ماشین می‌گوید که برای کسب اطلاعات بیشتر به این صفحه اشاره شده برود و یا دیتاست foaf که ارتباطات انسانی را تعریف می‌کند مانند شناختن، نام مستعار، پست الکترونیکی و از این قبیل. همانطور که می‌بینید چهار قانون مدنظر تیم برنرز-لی در شکل (۱-۱) رعایت شده به طوری که هم از یو-آر-آی استفاده شده، هم سایر لینک‌ها را در آن لحاظ شده و هم از زبان استانداردی مانند اُ-دبلیو-ال استفاده شده و در آخر از پروتکل اچ-تی-تی-پی جهت ارتباط استفاده شده است. بنابراین با اعمال چنین مکانیزمی درمورد کدهای برنامه‌نویسی نیز می‌توان زمینه را برای جستجوی بهتر کدها در فضای مجازی وب موثر کرد. اما