

مروری بر قالب بندی ایکس.ام.ال و کاربرد آن در بازیابی اطلاعات

محمدباقر دستغیب^۱

چکیده

با پیشرفت فناوری و فنون اطلاعات و ارتباطات در قرن اخیر، اهمیت دسترسی به اطلاعات صحیح و جامع روز به روز بیشتر می شود طبق آمارهای موجود، روزانه حدود ۷ میلیون صفحه جدید به فضای شبکه جهانی اضافه می شود. با توجه به حجم اطلاعات موجود در اینترنت، و ناهمگن بودن اطلاعات، لزوم به کارگیری روش های خودکار در نمایه سازی و بازیابی دقیق اطلاعات بیشتر احساس می شود. کاربران پرسش های خود را به نظام های بازیابی اطلاعات (موتورهای کاوش) ارائه می دهند و جواب نظام، باید مرتبط با پرسش کاربر باشد، ولی به دلیل حجم اطلاعات موجود در وب و نبود قالب مشخص که ساختار معنایی مدارک را حفظ نماید، دقت موتورهای کاوش کم می شود.

ساختار ایکس.ام.ال^۲ قالب مناسبی برای حفظ ساختار معنایی مدارک در سطح وب است. این قالب علاوه بر حفظ ساختار معنایی، دارای امکاناتی برای ساخت نمایش وبی مدارک در مرورگرهاست.

موتورهای کاوش بیشتر و کارایی بالاتری را در جستجوی مدارک در قالب ایکس.ام.ال از خود نشان می دهند. جستجوگرهای مدارک ایکس.ام.ال به سه دسته تقسیم می شوند: جستجوگرهایی براساس پایگاه داده ها، جستجوگرهایی براساس روش های بازیابی اطلاعات، و جستجوگرهای ترکیبی از انواع جستجوگرهای مدارک ایکس.ام.ال.

کلیدواژه ها

بازیابی اطلاعات، موتور جستجو، قالب بندی، ایکس.ام.ال

۱. مقدمه

با پیشرفت فناوری و فنون اطلاعات و ارتباطات در قرن اخیر، اهمیت دسترسی به اطلاعات صحیح و

۱. کارشناس ارشد مهندسی کامپیوتر کتابخانه منطقه ای علوم و تکنولوژی شیراز dstghaib@srst.com

2. XML = EXtensible Markup Language

جامع، روز به روز بیشتر می‌شود. تا سال ۲۰۰۱ تعداد صفحات^۱ شمارش شده در شبکه جهانی اینترنت ۴ میلیارد اعلام شده است، و روزانه حدود ۷ میلیون صفحه جدید به آن اضافه می‌گردد. با توجه به حجم انبوه اطلاعات موجود در اینترنت و ناهمگن بودن اطلاعات (طبق آمار ۲۰۰۱ م. ۸۳ درصد از صفحات تجاری، ۶ درصد علمی، و بقیه در سایر زمینه‌هاست)، تفکیک اطلاعات از یکدیگر، و به دست آوردن اطلاعات مناسب با نیاز کاربر، برای موتورهای کاوش^۲ امری دشوار است. قالب بیشتر صفحات در اینترنت، ابرمتن^۳ است، که قالب استاندارد مرورگرهای وب است. در این قالب بندی، مدارک با برچسب^۴های معنادار برای مرورگر، معرفی می‌شوند. در این قالب بندی، مدرک می‌تواند شامل متن، صوت و تصویر باشد. در قالب بندی ابرمتن، فقط قالب صفحه نمایش وبی مشخص می‌گردد و نشانه‌ها و معنای اشیاء^۵ مدل شده، مشخص نمی‌شود. در قالب بندی ابرمتن حدود ۱۰۰ تا ۱۲۰ برچسب از قبل تعریف شده، برای مدل سازی مدارک وجود دارد. این تعداد برچسب برای بیان مفهوم و معنای مدارک کافی نیست (۳).

مثال ۱-۱: سند ابرمتن برای ایجاد نمایش وبی یک مقاله

```
< HTML >
< TITLE > عنوان صفحه وب در مرورگرها < /TITLE >
< BODY >
< H1 > عنوان مقاله < /H1 >
< H4 >
چکیده مقاله
< /H4 >
< P >
متن مقاله
< /P >
< /BODY >
< /HTML >
```

همان طور که در مثال ۱-۱ مشاهده می‌شود، نوع مدرک مشخص نشده است، بنابراین کاوشگر برای تشخیص نوع مدرک، با مشکل مواجه خواهد شد. همچنین تشخیص کلیدواژه‌ها^۶ به سادگی میسر نخواهد بود. باید

دقت کرد، این مدرک دارای قالب بندی ساخت یافته بوده است و پس از تبدیل به قالب بندی ابرمتن، قالب اولیه اش را از دست داده است. برچسب های استفاده شده مانند <H1> و <H4>، فقط قالب نمایشی صفحه وب را مشخص می‌کنند و قالب معنایی اجزای مقاله را نمایان نمی‌سازند.

با توجه به مطالب فوق، به قالب بندی استاندارد نیاز است که علاوه بر مشخص کردن ساختار صفحه نمایش مدرک، ساختار معنایی مدرک را نیز بیان کند. در این صورت موتورهای کاوش به راحتی می‌توانند نوع مدارک و کلیدواژه‌های آن را تشخیص دهند و در صورت نیاز مدرک را بازیابی نمایند.

در ادامه، ابتدا درباره ساختار اسناد ایکس.ام.ال صحبت می‌شود و سپس روش های گوناگون کاوش در اسناد ایکس.ام.ال بررسی می‌گردد.

۲. قالب بندی ایکس.ام.ال برای صفحات وب

۲-۱. مروری بر ایکس.ام.ال

همان طور که بیان شد، نیاز به قالب بندی استاندارد، به منظور بیان قالب معنایی و نمایش مدرک است. قالب ایکس.ام.ال علاوه بر مشخص کردن ظاهر صفحه نمایش، معنای مدرک را نیز مدل می‌کند. استاندارد ایکس.ام.ال توسط ائتلاف شبکه جهانی^۷ معرفی شده است. این قالب بندی مانند ابرمتن از دسته زبان های نشانه گذاری^۸ است. در این زبان برچسب ها، معانی آنها، و شکل صحیح به کارگیری آنها، قابل تعریف است.

همان طور که در مثال ۱-۲ مشاهده می‌شود، در این قالب بندی علاوه بر مشخص شدن قالب نمایشی وبی مقاله (برای مثال عنوان مقاله بزرگ تراز متن مقاله است و در وسط تراز می‌گردد)، نوع مدرک (مقاله بودن با برچسب ARTICLE) و همچنین اجزای آن به تفکیک مشخص شده است. بنابراین کاوشگر وب به راحتی نوع مدرک و اجزای آن را تشخیص می‌دهد. باید توجه داشت این نوع قالب بندی برای مقاله، قبل از نمایش وبی موجود بوده و ایکس.ام.ال قالب معنایی مدرک را حفظ نموده است (۳؛ ۵).

در اسناد ایکس.ام.ال قالب اشیاء بهتر نمایان می‌شوند، بنابراین کاوشگر می‌تواند معنای اجزای درونی

1. Web Page

2. Search Engine

3. HTML = Hyper Text Markup Language

4. Browser

5. TAG

6. Objects

7. Key Words

8. W3C به نشانی www.w3c.com

9. Markup

یک شیء را درک نماید و سپس از آن استفاده کند. در مثال ۲-۲ ساختار تاریخ با قالب‌های گوناگون بیان شده است، اگر قالب بیان مدرک صحیح باشد، آن‌گاه درک معنای مدرک و دسترسی به اجزای آن بسیار ساده است. برای نمونه در مثال ۲-۲ قسمت ب، دسترسی به ماه و یا روز، بسیار ساده است (۵).

```

مثال ۱-۲: بیان قالب‌بندی یک مقاله با ایکس.ام.ال
<? Xml version = "1.0"? >
document>
  <article type = "iso">
    <title> عنوان مقاله </title>
    <author> نویسنده </author>
    <keywords> کلیدواژه‌های مقاله </keywords>
    <abstract> چکیده </abstract>
    <content> محتوای مقاله </content>
    <section no = "1"> فصل یک </section>
    <section no = "2"> فصل دو </section>
  </content>
</article>
</document>

```

مثال ۲-۲: تاریخ در ایکس.ام.ال

```

<data>
  4 December, 2003
</date>

```

الف

```

<date>
  <year> 2003 </year>
  <month> December</month>
  <day> 4 </day>
</date>

```

ب

۲-۲. ایکس.ام.ال و قابلیت‌های آن

دلایل زیادی برای محبوبیت ایکس.ام.ال نسبت به ابرمتن وجود دارد. در ادامه مهم‌ترین قابلیت‌های مدارک ایکس.ام.ال بررسی می‌شود.

– تبادل داده‌های آسان

در دنیای رایانه و در شبکه‌های رایانه‌ای مانند اینترنت، یکی از مهم‌ترین قابلیت‌ها، نیاز به تبادل داده‌ها میان نرم‌افزارهای گوناگون است. با توجه به متنی^۱ بودن مدارک ایکس.ام.ال و خودتوصیف بودن آن، امکان تبادل آن میان نرم‌افزارها به راحتی امکان‌پذیر است.

ساختار ایکس.ام.ال قالب مناسبی برای حفظ ساختار معنایی مدارک در سطح وب است. این قالب علاوه بر حفظ ساختار معنایی، دارای امکاناتی برای ساخت نمایش وبی مدارک در مرورگرهاست.

در مدارک ایکس.ام.ال، داده‌ها و برچسب‌ها به صورت متنی ذخیره می‌شوند، بنابراین پیکربندی^۳ آن به راحتی امکان‌پذیر است. پیکربندی برچسب‌ها و شخصی‌سازی^۴ آن زبان جدیدی ایجاد خواهد کرد که برای عموم قابل استفاده باشد. برای مثال زبان‌هایی برای نمایش و مدل‌سازی فرمول‌های ریاضی^۵ و یا نظام‌های بانکی^۶ پیوسته موجود است (۵؛ ۶).

– خودتوصیف بودن داده‌ها

در مدارک ایکس.ام.ال داده‌ها خودتوصیف است. برای نمونه در مثال ۲-۲ متوجه می‌شویم که یک تاریخ، با اجزای آن (سال و ماه و روز) بیان شده است و نیاز به توضیحی اضافه برای درک محتوای مدرک نیست. همچنین کاوشگر وب نیز از مزایای خودتوصیف بودن داده‌ها بهره می‌برد. در مثال ۲-۳ مبلغ یک جنس به صورت خودتوصیف بیان شده است (۵؛ ۶).

```

<LASTNAME>Dastgheib </LASTNAME>
  </NAME>
  <ID>2345ab</ID>
</PERSON>
<PERSON>
  <NAME>
    <FIRSTNAME> Ramin </FIRSTNAME>
    <LASTNAME> Rasti </LASTNAME>
  </NAME>
  <ID> 4871 sa </ID>
</PERSON>
</DOCUMENT>

```

همان طور که در مثال ۴-۲ مشاهده می‌گردد، ساختار معنایی مدرک در برچسب DOCTYPE مشخص شده است. بنابراین مدرک مندرج در مثال، خوش ساخت است. برای تعیین معتبر بودن مدرک باید قوانین معنایی موجود برقرار باشد. علامت * بعد از PERSON نشان می‌دهد که مدرک می‌تواند از صفر تا چند PERSON داشته باشد. پس از آن، ساختار معنایی برچسب PERSON مشخص شده است. این برچسب از دو قسمت نام (NAME) و آی.دی (ID) تشکیل شده است. همچنین برچسب نام از دو قسمت نام اول (FIRST NAME) و نام فامیل (LAST NAME) تشکیل شده است. معنای #PCDATA، قالب متنی برای داده‌هاست.

برای تولید مدارک ایکس.ام.ال و ویرایشگر^۷های مختلفی وجود دارد. با کمک این ویرایشگرها، می‌توان به راحتی مدارک ایکس.ام.ال را تولید کرد (۵؛ ۶).

۳. کاوشگر وب برای مدارک ایکس.ام.ال

۳ - ۱. مقدمه

یک کاوشگر وب، مانند آلتاویستا یا گوگل، هزاران مدرک را در پاسخ به یک درخواست بر می‌گرداند، که برخی از آنها مرتبط و برخی دیگر نامرتبط هستند. کاربر^۸ باید اسناد مرتبط را از غیر مرتبط جدا کند، این جداسازی، برای کاربر، عمل ساده‌ای نیست. با توجه به اطلاعات اضافه‌ای که علاوه بر داده‌ها در مدرک ایکس.ام.ال ذخیره می‌گردد، نوید تازه‌ای برای جستجوی دقیق‌تر وجود دارد.

مثال ۳-۲: خودتوصیف بودن داده‌ها در ایکس.ام.ال و بیان مبلغ یک جنس

```

<Item>
  <Name> نام </Name>
  <MadeIn> سازنده </MadeIn>
  <Price> قیمت </Price>
  <TechnicalInfo> اطلاعات فنی </TechnicalInfo>
</Item>

```

— ساخت یافتگی^۱ و جامعیت^۲ داده‌ها

در مدارک ایکس.ام.ال این قابلیت وجود دارد که علاوه بر تعریف داده‌ها، ساختار آن را نیز بیان کنیم. به عبارت دیگر، یک سری قواعد معنایی^۳ برای برچسب‌ها تعریف می‌شود، تا علاوه بر نمایش ظاهری ساختار، شکل صحیح به‌کارگیری آن را نیز مشخص کند. برای کنترل معتبر^۴ بودن ساختار مدارک ایکس.ام.ال، ابزارهایی^۵ وجود دارد. این ابزارها، مدرک را با قواعد معنایی آن تطبیق می‌دهند. مدرکی که قواعد ساختاری آن تعریف شده و همراه مدرک آورده شده باشد، مدرک خوش ساخت^۶ نامیده می‌شود. در مثال ۴-۲ یک مدرک خوش ساخت ایکس.ام.ال آورده شده است. در این مثال علاوه بر بیان داده‌ها، قواعد معنایی داده‌ها نیز آورده شده است. مثال ۴-۲: مدرک خوش ساخت

```

<?xml version="1.0"?>
<!DOCTYPE DOCUMENT>
  <!ELEMENT DOCUMENT(PERSON)*>
  <!ELEMENT PERSON(NAME,ID)>
  <!ELEMENT NAME (FIRSTNAME, LASTNAME)>
  <!ELEMENT FIRSTNAME (#PCDATA)>
  <!ELEMENT LASTNAME (#PCDATA)>
  <!ELEMENT ID (#PCDATA)>
  >
<DOCUMENT>
  <PERSON>
    <NAME>
      <FIRSTNAME> Mohammad </FIRS
      TNAME>

```

در XML به این قواعد (Document Type Definition) DTD گفته می‌شود. 3. Semantic Rules
 1. Structured 2. Integrity 4. Valid 5. Tools 6. Well Formed 7. Editor 8. Human User

این مدارک نسبت به مدل‌های قبلی مثل پی.دی.اف^۱ و یا ابرمتن با سرعت و دقت بهتری ذخیره و بازیابی می‌گردد. در حال حاضر الگوهایی از موتورهای کاوش وب، پیاده‌سازی شده است (۷) ولی تاکنون جنبه تحقیقاتی داشته است و به‌طور عملی توسط کاوشگرهای وب استفاده نشده است.

در ادامه سه روش مختلف برای ایجاد یک موتور کاوش وبی، به‌صورت اجمالی بررسی می‌شود. ابتدا روشی وابسته به پایگاه‌های داده^۲ بررسی می‌شود سپس، روشی مبتنی بر سازوکارهای بازیابی اطلاعات بررسی می‌گردد و در پایان روش ترکیبی مرور می‌گردد.

۳-۲. روش‌های وابسته به پایگاه داده‌ها

قالب بندی ایکس.ام.ال می‌تواند به راحتی بر داده‌های موجود در پایگاه داده‌ها اعمال گردد. برای مثال W3C، الگویی برای اعمال قواعد ایکس.ام.ال بر داده‌های پایگاه داده‌ها بیان کرده است (۲) که نشان می‌دهد چگونه داده‌های موجود در پایگاه داده رابطه‌ای^۳ می‌تواند به یک مدرک ایکس.ام.ال تبدیل گردد. هدف اصلی امکان پذیر ساختن نشر مدارک با پیکربندی استاندارد است. W3C برای این عقیده است که اطلاعات موجود در بانک‌های اطلاعاتی ابتدا به صورت ابرمتن منتشر می‌شود، و هم‌اکنون زمان انتشار آن اطلاعات با قالب بندی ایکس.ام.ال است.

تخمین زده می‌شود که حدود ۷۵ درصد از اطلاعات موجود در شبکه جهانی، از پایگاه‌های داده به دست می‌آید. به جای تبدیل این اطلاعات به مدارک ابرمتن، می‌توان اطلاعات موجود در پایگاه داده‌ها را به اسناد ایکس.ام.ال تبدیل کرد. در این صورت رابطه میان داده‌ها، به صورت خودتوصیف در مدارک ایکس.ام.ال حفظ می‌شود (۸).

مدارکی که از پایگاه داده‌های متمایز به دست می‌آیند، ممکن است دارای ساختار داخلی متفاوت باشند. بنابراین تبدیل اطلاعات موجود در پایگاه داده‌ها به مدرک ایکس.ام.ال بسیار ساده است ولی تبدیل مدرک ایکس.ام.ال به پایگاه داده مستوجب انجام عملیات خاص بر روی مدارک است. بنابراین یک ابرپایگاه داده^۴ مورد نیاز است تا بتواند

مدرک ایکس.ام.ال را به یک پایگاه داده القا نماید. بنابراین درخواستی که به یک ابرپایگاه داده می‌رسد، بین مجموعه‌ای از پایگاه‌های داده منتشر می‌شود، اگر ساختارهای موجود در ابرپایگاه نتواند ساختار داده‌ها را توصیف نماید، هیچ مدرکی تولید نخواهد شد. به کمک این سازوکار، از نبود دقت موجود کاسته می‌شود (۱).

۳-۳. روش‌های مبتنی بر بازیابی اطلاعات

سازوکارهای بازیابی اطلاعات، مستقیماً می‌تواند بر مدرک ایکس.ام.ال اعمال شود. هر مدرک ایکس.ام.ال

مدارک ایکس.ام.ال علاوه بر آنکه ساختار نمایشی وبی برای مدارک را نگهداری می‌کنند، اطلاعات ساختار معنایی مدارک را نیز ذخیره می‌کنند. بنابراین، پیکربندی معنایی سند پس از انتقال به فضای وب، حفظ می‌شود.

را می‌توان به عنوان یک مدرک متنی در نظر گرفت، و آن را پردازش کرد. نحوه برخورد نظام بازیابی اطلاعات با برچسب‌ها، می‌تواند متفاوت باشد. در ساده‌ترین حالت می‌توان برچسب‌ها را نادیده گرفت، مزیت این روش سادگی آن است، ولی باعث گم‌شدگی ساختار معنایی اطلاعات می‌گردد. در روش دیگر، برای ساخت فایل نمایه^۵، از برچسب‌های موجود در مدرک استفاده می‌شود، زیرا برچسب‌ها حاوی اطلاعات مهمی درباره ساختار معنایی و متن‌های موجود در مدرک هستند (۴). بسیاری از موتورهای جستجو، برای جستجوی یک کلیدواژه خاص، در به‌کارگیری برچسب‌ها رفتار گوناگونی نشان می‌دهند. برای مثال جستجوگر GOXml.com برچسب‌های واقع در درون یک برچسب را به عنوان مجاورت در نظر می‌گیرد. برای مثال، کاربر برای جستجوی «همایش فناوری اطلاعات»، می‌تواند پرسش خود را به صورت یک همایش بیان کند که حاوی فناوری اطلاعات باشد. این مثال کاربرد برچسب‌های تودرتو^۶ را در بازیابی اطلاعات نشان می‌دهد. تفاوت فاصله میان برچسب‌ها و کلیدواژه‌ها می‌تواند میزان

ارتباط میان مدرک و پرسش را بیان کند. بنابراین با کمک فاصلهٔ برچسب‌ها، می‌توان مدارک را ارزش‌گذاری نمود(۴؛۶).

۳-۴. روش‌های ترکیبی

برای جستجوی مجاورت کلیدواژه‌ها، روش دقیق‌تر آن است که برچسب‌های تودرتو مشخص شود. جستجوگر GOXml.com میان‌کلماتی که مستقیماً در متن برچسب به‌کار می‌روند، و کلمات موجود در متن درون برچسب تفاوت می‌گذارد. با کمک ارتباط و مسیرهای^۱ میان کلمات موجود در مدرک، می‌توان ارتباطات میان کلیدواژه‌ها را به‌دست آورد(۹). در روش ترکیبی ابتدا مدارک با کمک روش‌های بازیابی اطلاعات، نمایه‌سازی می‌شود، و در ذخیرهٔ اطلاعات و همچنین پرسش‌هایی که باید به‌نظام داده شود، از پایگاه داده‌ها استفاده می‌شود. به‌عبارت دیگر پرسش‌های کاربر به پرسش‌های ساخت‌یافته و استاندارد پایگاه داده‌ها، تبدیل می‌شود.

مزیت این روش آن است که با سرعت بالا می‌تواند پایگاه‌هایی را که باید جستجو شوند، مشخص می‌کند و بیان درخواست و پرسش کاربر با زبان استاندارد پایگاه داده‌ها، باعث کارایی بهتر نظام بازیابی از نظر زمان و فضای دیسک می‌شود، برای مثال XYZFind نظامی است که با دریافت مدارک ایکس.ام.ال، نمایه‌سازی خودکار را انجام می‌دهد و پرسش‌ها را استخراج می‌نماید. در این نظام از سازوکارهای بازیابی اطلاعات و روش‌هایی بر پایهٔ پایگاه داده‌ها به‌طور ترکیبی استفاده شده است(۱۰).

نتیجه‌گیری

کاوشرهایی که مدارک ایکس.ام.ال را نمایه‌سازی می‌کنند، نوید جستجوی دقیق‌تر در انبوهی از مدارک غیرهمگون را به ارمغان می‌آورند. مدارک ایکس.ام.ال علاوه بر آنکه ساختار نمایشی و بی برای مدارک را نگهداری می‌کنند، اطلاعات ساختار معنایی مدارک را نیز ذخیره می‌کنند. بنابراین، پیکربندی معنایی سند پس از انتقال به فضای وب، حفظ می‌شود. کاوشرهایی ایکس.ام.ال با سه روش، عملیات

بازیابی را مدل می‌کنند. روش اول، از تکنیک پایگاه داده‌ها استفاده می‌کند. مزیت این روش حفظ جامعیت و رابطهٔ میان داده‌ها، استفاده از زبان استاندارد پرسش برای پرسش‌های کاربر، و ارتباطات استاندارد میان اشیاء^۲ داده‌ای در پایگاه داده‌هاست.

روش دوم مبتنی بر سازوکارهای بازیابی اطلاعات است. در این روش از راه‌کارهای نظام‌های کلاسیک بازیابی اطلاعات استفاده می‌شود. مزایای این روش، علاوه بر مزایای موجود در مدارک ایکس.ام.ال، امکان به‌کارگیری سازوکارهای بازیابی اطلاعات مانند فهرست‌های معکوس در بازیابی اطلاعات است.

روش سوم، ترکیب روش اول و دوم است، بنابراین علاوه بر حفظ ساختار معنایی داده‌ها، جامعیت داده‌ها و رابطهٔ میان آنها، کارایی نظام نیز در سطح قابل قبولی قرار دارد. به‌طور کلی استفاده از ایکس.ام.ال به‌مثابهٔ قالب استاندارد برای بیان مدارک در سطح وب کمک می‌کند که ساختار معنایی مدارک حفظ گردد. علاوه بر آن قابلیت‌های موجود، برای ساختار نمایشی مدارک در وب نیز استفاده شود، بنابراین پیشنهاد می‌گردد با توجه به حجم کم، قانونمند بودن، و ساختار مناسب ایکس.ام.ال، از ساختار ایکس.ام.ال برای قالب آیندهٔ مدارک اینترنتی استفاده گردد.

منابع

1. Abiteboul, S. *On views and XML*. [S.L]: SIGMOD Record, 1999.
2. Bos, B. "XML representation of a relational database". 1997. [on-line]. Available: <http://www.w3.org/XML/RDB.html>.
3. Egnor, D.; Lord, R. *Structured Information Retrieval Using XML*. USA: XYZFind, 2000.
4. Frakes, W.; Baeza-Yates, R. *Information Retrieval: Data Structures and Algorithms*. [S.L]: Prentice-Hall, 1992.
5. Holzner, H. *Inside XML*. [S.L]: New Riders Publishing, 2001.
6. Kushmerich, N. *Expressive Retrieval from XML documents*. Ireland: University of Dublin, 2000.
7. Navarro, G.; Baeza-Yates, R. "Proximal nodes: a model to query document databases by content and structure". *ACM Trans. on Information*

1. XPath
2. SQL = Standard Query Language
3. Object

Systems, Vol.15, No.4(1997).

8. Turau, V. "Making legacy data accessible for XML applications". 1999. [on-line]. Available: <http://citeseer.nj.nec.com/turau99making.html>.

9. W3C. "XML Path Language". <http://www.w3.org/TR/xpath>, 1999.

10. XYZFind Corporation. "XYZFind white paper: a new technology for search over heterogeneous structured data". 2000. [on-line]. Available: http://www.xyzfind.com/paper_001.pdf.

تاریخ دریافت: ۱۳۸۲/۱۰/۱۴