

جدول راهنما: روشی سه مرحله ای برای پردازش تفکیکی برگ پرس و جویهای XML

وحید گرکانی، جواد نوروزی و مصطفی حق جو سانجی

دانشگاه علم و صنعت ایران - دانشکده مهندسی کامپیوتر

{vahidgarakani, j.norози, haghjoom}@comp.iust.ac.ir

چکیده - با فراگیر شدن کاربرد XML در وب، تعداد و حجم سندهای XML افزایش یافته است. در نتیجه اهمیت پردازش بهینه پرس و جویهای XML بیشتر از قبل به چشم می خورد. پرس و جویهای XML شامل مجموعه ای از المانها می باشد که تحت یک ساختار درختی به هم مربوط هستند. روش هایی که برای پردازش پرس و جویها تاکنون مطرح شده است را می توان به دو دسته تقسیم کرد: (۱) روش هایی که با اعمال مستقیم پرس و جو در سند، به طور کورکورانه، سعی در رسیدن به جواب پرس و جو دارند (۲) روش هایی که با استفاده از شاخص های مسیری سعی می کنند محدوده جستجو در سند را برای پرس و جویهای تک شاخه ای کوچک کنند. این دو روش هر کدام تنها برای پرس و جویهای خاصی کارایی لازم را دارند و برای دسته بزرگی از پرس و جویها زمان پردازشی بالایی دارند. در این مقاله با ارائه ایده جدول راهنما، روش جدیدی را برای پردازش پرس و جوی XML که ترکیبی از دو روش قبلی است، مطرح شده است. جدول راهنما، گره های پردازش شونده، نحوه پردازش آنها و حتی نحوه رسیدن به جواب پرس و جو را نیز نشان می دهد. به طوری که به عنوان یک راهنما، ما را از پردازش کورکورانه در سند نجات می دهد. کارایی این روش با استفاده از محک XMark, TreeBank, DBLP در مقایسه با روش های مطرح در این زمینه نشان داده شده است.

کلید واژه - انطباق شاخه ای، جدول راهنما، شاخص های مسیری، بهینه سازی پرس و جو

۱- مقدمه

را دارد، می باشد. روش های زیادی برای برآورده کردن این هدف ارائه شده است که ما در زیر آنها را به ۴ گروه تقسیم کرده ایم:

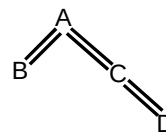
- گروه الف: *Structural Join* اولین روشی است که برای پردازش پرس و جویهای XML ارائه شده است. در این روش یک پرس و جو به تعدادی رابطه پدر-فرزند و یا جد-نسل تبدیل می شود. برای مثال Q1 به ۳ رابطه جد-نسل، A//B و A//C تبدیل می شود و در مرحله بعد ادغام دودویی صورت می گیرد. مشکل این روش، تولید زیاد داده های میانی است [۱].

- گروه ب: اولین بار در [۲] holistic twig join ارائه شد. این روش سعی می کند حجم داده های میانی را کم کند و عمل شکستن پرس و جو را با استفاده از دو روش path stack و twig stack انجام ندهد. بهینه سازی های زیادی روی این روش انجام گرفته است. برخی مقالات این روش را

با افزایش تعداد و حجم سندهای XML در وب، اهمیت پردازش کارای پرس و جویهای XML بیشتر به چشم می خورد. سندهای XML ذاتا درختی (گرافی شکل) هستند. خوشبختانه پرس و جویهای XML را نیز می توان با یک ساختار درختی بنام Query Tree Pattern (QTP) نشان داد. برای مثال پرس و جوی Q1 را در نظر بگیرید:

Q1: //A[./B]//C//D;

ساختار درختی Q1 در شکل (۱) آمده است. هدف تمامی روش های پردازش پرس و جو، پیدا کردن تمامی نمونه QTP های پرس و جو در سند واقعی، که حجم بسیار بالایی



شکل ۱: ساختار درختی پرس و جو Q1