



ارائه شده توسط:

سایت ترجمه فا

مرجع جدیدترین مقالات ترجمه شده

از نشریات معتبر

یک روش دارای نمونه های هدایت شده برای تغییرات آرام در پایگاه

ساختار بندی افزایش اطلاعات وب

چکیده : تغییرات آرام در پایگاه اطلاعاتی وب یک راه حل موثر برای ترکیب اطلاعات عمیق در وب می باشد . تا جائیکه به اطلاعات ما مربوط می شود ، روشهای موجود فقط متمرکز بر چگونگی تغییرات آرام تمام گزارشات در یک پایگاه اطلاعات وب می باشد . با توجه به دینامیک زیاد در پایگاه اطلاعاتی وب ، برداشت حجم کم گزارشات جدید با ایجاد تغییرات آرام در پایگاه اطلاعاتی علمی نمی باشد . این تحقیق مسئله تغییرات آرام در پایگاه اطلاعاتی وب را بررسی می کند که هدف آن بررسی موثر تغییرات آرام در گزارشات جدید یک پایگاه اطلاعاتی وب می باشد . در روش مطرح شده ، یک مدل نموداری جدید که یک نمودار جستجو است برای انتقال وظیفه افزایش تغییرات در یک فرایند وارونه شده است . براساس این نمودار ، جستجوهای مناسب برای تغییرات ایجاد می شود که با نمونه های گرفته شده از پایگاه اطلاعات وب هدایت می شوند . ارزیابی های تجربی گسترده از پایگاه اطلاعات وب ، کارایی تکنیک های مورد استفاده ما را تایید کرده و اطلاعاتی را در مورد تلاشهای آینده در این جهت فراهم می آورد .

لغات کلیدی : پایگاه اطلاعاتی وب - ترکیب داده های عمیق وب - تغییرات آرام در پایگاه

اطلاعاتی وب

1-مقدمه

وب عمیق اشاره به داده های در پایگاه اطلاعاتی وب دارد که محتوای بسیاری از آنها در مورد گزارشات اطلاعات ساختار بندی شده است . اعتقاد براین است که این مورد بزرگترین منبع

وجود اطلاعات وب است و از اینراه حل جمع آوری گزارشات ساختار بندی شده از پایگاه اطلاعات وب برای انجام تحقیق و پردازش اطلاعات به صورت مرکزی وب عمیق است . توسعه سریع سخت افزار کامپیوتری و اینترنت باعث شده است تا این جواب کاربردی تر از قبل شود . تلاشهای قبلی انجام شده فقط تاکید بر تغییرات آرام در پایگاه اطلاعاتی داشته اند و هدف به حداکثر رساندن پوشش اطلاعات پایگاه های وب است . ما این روش را تغییرات گسترده نامیده ایم . همان طور که مشخص است ، بسیاری از پایگاه های اطلاعاتی بسیار پویا هستند . برای مثال گزارشات جدید به صورت پیوسته وارد می شوند . برای اطمینان از ثبات اطلاعات ترکیبی وب ، استفاده از روش تغییرات آرام برای برداشت حجم کم گزارشات جدید همیشه کاربردی نیست و منجر می شود تا حجم زیادی اطلاعات غیر لازم در وب و شبکه ایجاد شود . در این تحقیق ، یک مسئله مهم و حل نشده را برای رسیدن به جواب همراه با تغییر مورد بررسی قرار می دهیم ، چگونه می توان به گزارشات جدید بدون ایجاد تغییر در کل پایگاه اطلاعاتی دست یافت ؟

برای این منظور یک روش افزایش تغییر همراه با نمونه های هدایت شده ارائه کرده ایم . ایده کلی این روش به صورت زیر توصیف شده است : اول اینکه تعداد کمی از نمونه های تصادفی از پایگاه اطلاعاتی وب گرفته شده اند . سپس با آنالیز انحراف بین نمونه ها و ویرایش تاریخی ، یک گزارش درست برای ایجاد جستجو های مناسب برای گزارشات جدید ارائه شده است . در این روش ، یک مدل نموداری جستجو محور ارائه کردیم و بنابراین اطلاعات داده شده می تواند به صورت یک نمودار بدون جهت بر پایه مدل نمایش داده شوند . وظیفه ایجاد تغییر فراینده تبدیل به یک فرآیند نموداری می شود که در آن تغییر از نمودار نمونه ها در پایگاه اطلاعات وب شروع می شود و در هر مرحله V انتخاب می شود و جستجوی مناسب با استفاده از بردار انتخاب شده برای تغییر انجام می شود . چون فقط یک روش جستجو برای

ارزیابی پایگاه داده های وب وجود دارد، ایجاد جستجوی اتوماتیک یکی از روشهای کلیدی است. هدف ما به حداکثر رساندن پوشش گزارشات جدید و به حداقل رساندن پوشش اطلاعات قدیمی به صورت همزمان است. همانند تحقیق اول برای بیان مسئله تغییرات آرام در پایگاه اطلاعاتی وب، تاکید این تحقیق هم به صورت زیر خلاصه شده است. ابتداء مسئله جدید تغییرات آرام در پایگاه اطلاعاتی وب را شناسایی کردیم. برخلاف تحقیقات گسترده انجام شده در مورد تغییرات وب، بیان کردیم که مسئله اصلی پایگاه های موثر در ثبات بین پایگاه اطلاعاتی بومی و پایگاه های مرکب وب است. ثانیاً، یک چهار چوب تئوری ارائه کردیم که تغییرات پایگاه اطلاعات را به صورت یک نمودار مدلسازی کرده ایم. در تفاوت مدل های نموداری مطرح شده توسط تحقیقات قبلی، مدل نموداری ما در سطح گزارشات است که می تواند توصیف کن که آیا هیچ دو گزارش جستجو با هم مرتبط هستند. سوم اینکه، بر اساس مدل نموداری، روشهای ساده و هوشمندی را مطرح کردیم که برای مسائل کلیدی در روش تغییرات فراینده به کار گرفته می شوند و هدف آنها ایجاد جستجوهای مهم برای برداشت گزارشات جدید است

ادامه این تحقیق به صورت زیر مرتب شده است: بخش 2 موارد اولیه را بیان کرده است. بخش 3 یک مدل نموداری جستجو محور را نشان می دهد. روش انتخاب جستجو بر پایه مدل نموداری در بخش 4 داده شده است. یافته های تجربی در بخش 5 مورد بحث قرار گرفته است. بخش 6 بعضی از تحقیقات مربوط را موروی می کند و بخش 7 هم نتیجه گیری تحقیق است.

- 2 مواد اولیه

- A نمونه های WDB

نمونه های WDB اشاره دارد به یک مجموعه گزارش که به صورت تصادفی از WDB انتخاب می شوند چون تنها راه کلی دسترسی به پایگاه اطلاعاتی وب از طریق سطح جستجو است ، تکنیک های نمونه گیری سنتی نمی توانند به آسانی به کار گرفته شوند زیرا دسترسی کامل به WDB وجود ندارد . خوشبختانه ، در بخش 17 ، یک روش قدم زنی تصادفی برای نمونه گیری اطلاعات وب مطرح شده است . در این روش ، تکذیب احتمالی برای ارتقای کیفیت نمونه استفاده می شود و اندازه نمونه ها می تواند با تغییر پارامترها تغییر کند . در این تحقیق ، ما فرض می کنیم که نمونه ها اخیرا گرفته شده اند . در حقیقت ، انتخاب بدون تعجب نمونه ها غیر ممکن است زیرا جستجو برای نمونه گیری با تعجب همراه است در آزمایشات ، ما در مورد تاثیر عملکرد هم بحث می کنیم که توسط نونه های تعصبی حاصل شده است .

نسخه تاریخی WDB

همان طور که قبلا بیان شده است ، دیپ وب بسیار متغییر است . ما از WDB برای اشاره به نسخه تاریخی WDB استفاده می کنیم . برای مثال ، گزارشات WDB که قبلا تغییر یافته بودند در پایگاه اطلاعاتی ذخیره شده بودند در این تحقیق ، ما فقط سناریوی مربوط به وارد کردن گزارش به حوزه های شغلی و تحقیقی را در نظر می گیریم و از WDBh برای اشاره به گزارشات جدید وارد شده در طول دوره استفاده می کنیم . نسخه تاریخی WDB می تواند به آسانی از پایگاه اطلاعاتی و با استفاده از SQL به دست آید . از این طریق ، نسخه تاریخی WDB می تواند یک چشم انداز مجازی برای پایگاه اطلاعاتی در نظر گرفته شود .

. C گسته سازی مقادیر مربوط به شاخص ها

در این تحقیق ، ما فرض می کنیم که مقدار شاخص های داده شده در سطح جستجو به صورت یک مجموعه از مقادیر گسته است . در حقیقت ، سه نوع انتخاب داده برای کاربر وجود دارد تا یک جستجو کامل شود : لغات کلیدی ، انتخاب و دامنه شاخص انتخاب توانسته است نیاز مربوط راتامین کند . در مورد شاخص بیان می شوند در نتیجه ، بیش از یک مقدار می تواند برای یک شاخص در یک گزارش یافت شود . برای مثال عنوان یک تحقیق دارای چند لغات کلیدی است . برای شاخص دامنه هم مقدار به مجموعه ای از دامنه های کوتاه تقسیم می شود از این طریق ، متمم شاخص ها می توانند بدون اختلاف پیش روند .

. عملکرد تغییرات افزایشی

هدف ایجاد تغبیر افزایشی در وب ، ایجاد در گزارشات جدید و به حداقل رساندن هزینه های ارتباطی است . در نتیجه ، دو عامل در نظر گرفته می شود : میزان پوشش و هزینه تغییرات مورد اول برای سنجش موثر بودن استفاده می شود در حالیکه مورد دوم برای سنجش کارایی استفاده می شود .

-مدل نموداری جستجو محور

. این مدل QRG که برای WDB یک نمودار بدون جهت است که می تواند به صورت زیر ساخته شود : برای هر گزارش r_i یک بردار v_{iev} وجود دارد. مقدار حاشیه برای fri و rj هم می تواند دارای یک شاخص q باشد که می تواند در سطح شاخص نمایش داده شود .

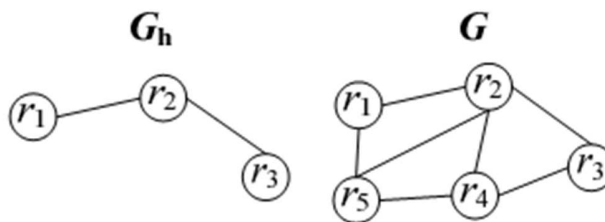
با توجه به تعریف این مدل ، پیچیدگی نمودار برای WDB داده شده توسط دو عامل مشخص می شود : تعداد گزارشات و قابلیت جستجو. در اینجا قابلیت جستجو اشاره دارد به تعداد شاخص های موجود در سطح جستجو و اندازه بزرگتر هر شاخص منجر به حاشیه بیشتر در نمودار می شود . یک ویژگی جالب QRG این است که اگر بردار V برای جستجو انتخاب شود ،

بردارهای مرتبط با گزارشات بیشتری برگشت می خورد ما از این ویژگی برای انتخاب بردار و ارزیابی تعداد گزارشات برگشت خورده استفاده می کنیم . با توصیف پایگاه اطلاعاتی ساختار بندی شده وب با استفاده از QRG ، یک وظیفه تغییر به صورت یک نمودار تبدیل می شود که در آن تغییرات نسخه تاریخی WDB شروع می شود و در هر مرحله بردار X برای شکل گیری شاخص و گزارشات جدید در مجاورت V استفاده می شود و اطلاعات برای تغییرات آینده ذخیره می شوند .

شکل 1 نشان دهنده یک مثال برای تشریح نمودار مطرح شده است . شکل 1 نشان دهنده WDB و WDBH است که هر کدام دارای دو شاخص می باشند . شکل (1) b هم نمودار جستجو برای دو مورد بالاست . انتخاب های متفاوتی بر r_4 و r_5 وجود دارد برای مثال ، یک انتخاب می تواند r_4 و استفاده از b_2 در B و r_5 و r_3 و a_3 در B برای تغییرات r_4 باشد بدیهی است که انتخاب های مختلف منجر به پوشش مختلف و هزینه می شود در بخش بعد ، روش خود را برای گزارشات جدید و QRG ارائه می کنیم .

WDB_h			WDB		
	A	B		A	B
r_1	a1	b1, b2	r_1	a1	b1, b2
r_2	a2	b2, b3	r_2	a2	b2, b3
r_3	a3	b3	r_3	a3	b3
			r_4	a3	b3, b4
			r_5	a2	b2, b5

(a)



(b)

روش تغییر در پایگاه اطلاعاتی وب

ما ابتدا روش مطرح شده را بررسی می کنیم. نظریه اصلی برای این روش تغییر در گزارشات جدید و آنالیز انحراف بین نمونه ها و نسخه تاریخی در پایگاه اطلاعاتی وب است. ما تصور می کنیم که نمونه های انتخاب شده بدون تعصب برای WDB در پایگاه های منطقه ای ذخیره شده اند. ما از S برای اشاره به نمونه ها استفاده می کنیم. معمولا مسئله ما می تواند به صورت زیر ترسیم شود: با بیان S و WDB_h ، چگونه می توان به گزارشات جدید عضو $WDB-WDB_h$ دست پیدا کرد؟

شکل 2 نشان دهنده یک بررسی از مدل مل می باشد که یک فرایند محوری است. عملکرد سه مولفه در هر دور به صورت زیر است:

Incremental Web Database Crawling Algorithm

Input: S, WDB_h //The samples of WDB and the history version of WDB

Output: NRS // the set to store the crawled new records

Begin

```

1  Produce the query-related graphs  $G_S$  and  $G_h$  for  $S$  and  $WDB_h$ 
   respectively;
2  While the terminate criteria is not met
3     $r$  = RecordSelection( $G_S, G_h$ );
4     $q$  = QueryGeneration( $r$ );
5     $RS$  = RecordExtraction( $q$ );
6     $NRS$ .Add( $RS$ );
7  Return  $NRS$ ;
```

End

انتخاب گزارش: یک گزارش از WDB_h انتخاب کنید.

تولید جستار: با استفاده از گزارش تولید شده یک جستار تولید کنید.

استخراج گزارش: گزارشات را از نتایج استخراج کنید.

انتخاب گزارش و تولید جستار از مولفه های کلیدی در حوزه های تحقیقی برای استخراج اطلاعات وب هستند که به صورت گسترده مورد بررسی قرار گرفته اند، بنابراین، هیچ بحثی

در این تحقیق آورده نشده است. معیار توقف فرایند تغییر به دست آوردن اطلاعات جدید از 10 جستار نمی باشد. در ادامه، تکنیک های اصلی برای انتخاب و تولید جستار معرفی شده اند.

B. انتخاب گزارش

وقتی که نمودارهای جستار برای G_s و G_h تولید می شوند، مسئله انتخاب گزارش مربوط به چگونگی انتخاب یک بردار از بردارهای انتخاب شده است. با توجه به تعریف PNR، بردار انتخاب شده در هر دور باعث شکل گیری نسبت بردارهای جدید از میان بردارهای مجاور در G می شود. به عبارت دیگر، فاصله از بردار انتخاب شده و بردارهای جدید، باید تا حد امکان نزدیک باشد و فاصله بردار منتخب از بردارهای قدیمی تا حد امکان دور باشد. به هر حال، G مشخص نیست و ما از G_s به عنوان نماینده G استفاده کرده و فرض می کنیم S نمونه های انتخاب شده بدون تعصب هستند. با توجه به این مورد، معیار انتخاب بردار با فرمول زیر نمایش داده می شود

رابطه 4

به طوری که $\text{shortest path}(v, G_h)$ اشاره به میانگین کوتاه ترین مسیر بین v و بردارهای G_h دارد و این احتمال وجود دارد که v به بردارهای جدید متصل باشد. با توجه به معادله 4، الگوریتم انتخاب گزارش در شکل 3 نشان داده شده است.

Record Selection Algorithm
Input: G_s, G_h Output: v // the selected vertex in $G_s \cup G_h$

Begin

1 For each vertex v in $G_s \cup G_h$ do2 Compute $Sel(v)$;3 $Sel_Sum = \sum Sel(v_i) \quad v_i \in G_s \cup G_h$ 4 For each vertex v in $G_s \cup G_h$ do5 $Prob(v) = \frac{Sel(v)}{ACT_Sum}$; //the probability of selecting v 6 Randomly select a vertex v from $G_s \cup G_h$ according to $Prob(v)$;7 Return v ;

End

C. ایجاد یک جستار

بعد از اینکه یک بردار (گزارش) انتخاب شد، یک جستار مناسب با استفاده از بردار انتخاب می‌کنیم. هر جستار می‌تواند به صورت مجموعه‌ای از شاخص‌های کیفیت پیش‌بینی شود. همانند کارهای کنونی انجام شده در مورد تغییرات پایگاه اطلاعاتی، تاکید ما بر ساده‌ترین جستارها می‌باشد. با بیان یک گزارش، چند جستار می‌تواند انتخاب شود. برای مثال، سه جستار $A=a1$ ، $B=b1$ و $B=b2$ می‌توانند با استفاده از $r1$ در شکل 1 ایجاد شوند.

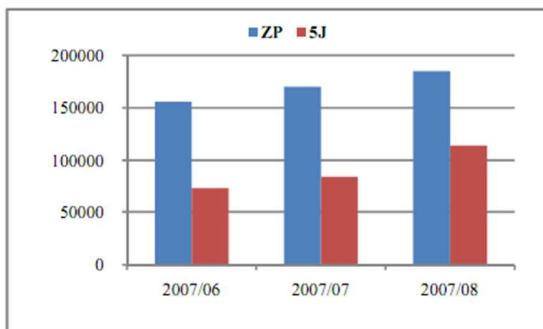
با در نظر گرفتن سنجش هزینه‌ها برای هر ICR ، هدف جستار انتخاب موثرترین جستار از میان جستارهای موجود است. کارایی جستار با استفاده از رابطه زیر تعریف می‌شود:

$$eff(q) = \frac{\text{count}(WDB, q) - \text{count}(WDB_h, q)}{\text{count}(WDB_h, q)} \quad (5)$$

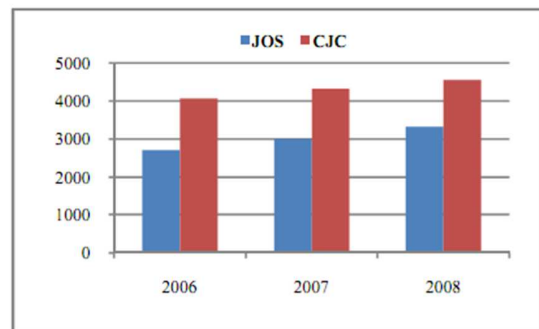
به طوری‌که $\text{count}(WDB, q)$ اشاره به تعداد گزارشات برگشت خورده در هنگام ارسال q به WDB می‌باشد. $\text{count}(WDB, q)$ نی‌تواند با استخراج عدد پنهان صفحه اصلی مثل "8503" به دست آید. با توجه به آمار برای پایگاه‌های اطلاعاتی بزرگ، بیش از 98.4 درصد از پایگاه‌های وب نشان‌دهنده اعداد پنهان در صفحات نتایج برای ثبت جستار می‌باشند. یک روش در قسمت 5 به کار گرفته شده است که برای استخراج اعداد پنهان استفاده می‌شوند).

count (WDB,q) می تواند با انتشار انتخاب شماره از WCBh به دست آید . در نتیجه
 $count (WDB,q) - count (WDBh,q)$ تعداد گزارشات جدید حاصل می باشد.

Web database	Home page	Queryable Attributes	Start time	End time
Zhaopin.com(ZP)	www.zhaopin.com	Position, Company, Location	2007/06/01	2007/08/31
51Job.com(5J)	www.51job.com			
Journal of software(JOS)	www.jos.org.cn	Title, Author, Key Words	2006/01/01	2008/12/31
Chinese Journal of Computers(CJC)	cjc.ict.ac.cn			



(a)



(b)

بر اساس معادله 5 ، یک روش ساده برای تولید جستار به صورت زیر توصیف می شود : اول اینکه ، تمام جستارها با کیفیت یک با استفاده از گزارش انتخاب شده تولید می شوند. ثانياً، کارایی را برای هر کدام از جستارها محاسبه کنید. در پایان، جستار دارای بیشترین کارایی را برای هر تغییر استفاده کنید.

۷. ارزیابی تجربی

در این تحقیق، ابتدا مجموعه اطلاعاتی استفاده شده در آزمایشات را ارائه کردیم ، سپس یک مجموعه از آزمایشات برای ارزیابی عملکرد اجرای تغییرات بر پایه روش مطرح شده انجام شده است.هم تغییرات و هم پایگاه اطلاعاتی توسط یک سرور ویندوز 2.66G با رم 3.25G و قابلیت اجرا در MySQL5RDBMS میزبانی شدند.

A. مجموعه اطلاعاتی

برای ارزیابی دقیق تغییرات آرام، مجموعه اطلاعاتی ما به صورت تاریخچه از چهار پایگاه اطلاعاتی واقعی و حوزه های تحقیق JOB TONG و C-DBLP فراهم می شوند. چهار پایگاه اطلاعاتی مورد استفاده عبارتند از زائوپین دات کام ZO، 51 جاب دات کام J، ژورنال اف سافتور JOS و مجله چینی کامپیوترها CJC. جزئیات این پایگاه اطلاعاتی در جدول 1 نشان داده شده است. در این جدول، ستون دوم نشان دهنده ویژگیهایی است که می تواند برای هر سطحی بررسی شوند. ستون های سوم و چهارم هم بیانگر تاریخچه آزمایشات انجام شده در طول دوره می باشد. شکل 4 نشان دهنده اندازه نسخه های تاریخچه است و بردار X اشاره به تعداد گزارشات و Y دلالت بر خط زمانی دارد. در حوزه شغلی، 3 نسخه برای هر کدام از پایگاه های اطلاعاتی وجود دارد و در حوزه تحقیق هم 3 نسخه موجود است. میانگین سرعت رشد در حوزه شغلی حدود 9 و 5 درصد در هر ماه است و میانگین رشد در حوزه شغلی 12.1 درصد در سال است.

B. آزمایشات انجام شده در مورد میزان پوشش و هزینه انتقال

ما عملکرد تغییرات را بر میزان پوشش CCR و ICR و هزینه تغییرات PNR را در چهار مجموعه اطلاعاتی بررسی کردیم. یک گروه آزمایش انجام شده است و جدول 2 نشان دهنده نتایج آنها می باشد. ما از آخرین نسخه برای تغییرات نسخه بعدی استفاده کردیم. برای مثال، در بالای ستون سوم در جدول 2 از نسخه 7 استفاده شده است. نسبت نمونه ها در هر کدام از پایگاه های اطلاعاتی 15 درصد تنظیم شده است و نمونه ها به طور تصادفی انتخاب شده اند. همانطور که با توجه به جدول 2 می توان مشاهده کرد، روش مبتنی بر تغییرات عملکرد بالایی برای میزان پوشش و هزینه ها دارد. برای CCR، تمام موارد بیش از 96 درصد و نزدیک به 100 می باشند. این مطلب به این معناست که متمم گزارشات در نسخه کنونی WDB می تواند مورد استفاده قرار بگیرد و فکر می کنیم دلیل این است که پایگاه اطلاعاتی

بزرگتر از مجموعه اطلاعاتی بررسی شده توسط محققان می باشند. برای PNR، تمام موارد بالا در یک دامنه ثابت هستند و این مطلب دلالت بر این دارد که 1. تغییرات بسیار منسجم هستند و 2. نسبت گزارشات جدید در تمام گزارشات ثابت است.

WDB	Measure	07	08
ZP	CCR	97.9%	99.1%
	ICR	86.3%	91.5%
	PNR	19.9%	26.2%
5J	CCR	97.2%	98.6%
	ICR	80.8%	91.3%
	PNR	28.5%	35.1%

(a) Job domain

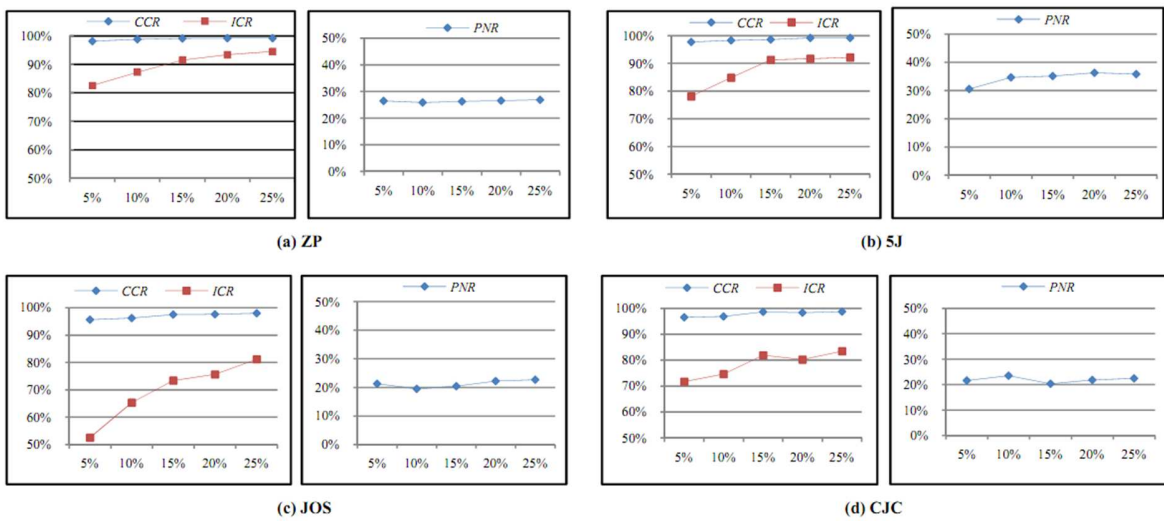
WDB	Measure	2007	2008
JOS	CCR	96.2%	97.5%
	ICR	64.0%	73.4%
	PNR	18.4%	20.4%
CJC	CCR	98.6%	98.7%
	ICR	80.7%	81.9%
	PNR	23.5%	24.6%

در مقایسه با روشهای مربوط به تغییرات گسترده، روش مطرح شده ما بهتر از روشهای دیگر برای سنجش CCR و PNR است. مورد 3 وضعیت کنونی تغییرات است. بهترین عملکرد برای CCR در نتایج تجربی حدود 90 درصد گزارش شده است که کمتر از مفدار میانگین است. با توجه به هزینه های گزارش شده، هزینه گزارشات برگشت خورده 1.2 تا 1.5 برابر گزارشات تغییرات است. اگر سرعت رشد برای WDB حدود 10 درصد باشد، ICR کمتر از 10 درصد است.

C. آزمایشات انجام شده در مورد اندازه های مختلف نمونه ها

در آخرین آزمایشی که انجام شده است، نسبت نمونه ها در هر کدام از پایگاه داده ها 15 درصد تنظیم شده است. برای ارزیابی تاثیر این نسبت بر عملکرد، این نسبت را به صورت مقادیر مختلف تنظیم کرده و آزمایشات مربوط به تغییرات آرام افزایشی را مجدد بررسی

کردیم. در حوزه شغلی هم ، هدف ایجاد تغییر رسیدن به نسخه های 80 است و در حوزه تحقیق هم هدف ایجاد تغییر رسیدن به نسخه 2008 است. شکل 5 نشان دهنده نتایج تجربی در مورد نسبت های مختلف است . دو نتیجه گیری می تواند بیان شود. اول اینکه ، اندازه نمونه ها تاثیر گذار بر CCR-ICR می باشد. در کل ، زمانی که نسبت بیش از 15 درصد باشد ، CCR و ICR با ثبات می باشند. بنابراین ، این نسبت می تواند در عمل 15 درصد تنظیم شود زیرا بدست آوردن اندازه بزرگ برای نمونه ها یک مسئولیت بزرگ است. ثانیاً، شواهدی وجود ندارد که بیان کند این نسبت بر PNR تاثیر می گذارد. دلیل هم این است که نسخه های بالاتر می توانند انتخاب را آسانتر کنند و این عامل می تواند نسبت گزارشات جدید را در گزارشات برگشت داده برای هر جستجو افزایش دهد.



D. آزمایشات انجام شده در مورد نمونه های انتخاب شده با تعصب

در عمل ، داشتن نمونه های واقعی غیر ممکن است به طوری که ارزیابی روش برای نمونه های انتخاب شده بدون تعصب کاری ضروری می باشد. نسبت هم 15 درصد تنظیم شده است. با توجه به نتایج به دست آمده در شکل 6، می توانیم متوجه شویم که عملکرد نهایی در نمونه های انتخاب شده بدون تعصب بهتر از نمونه های انتخاب شده با تعصب است. این

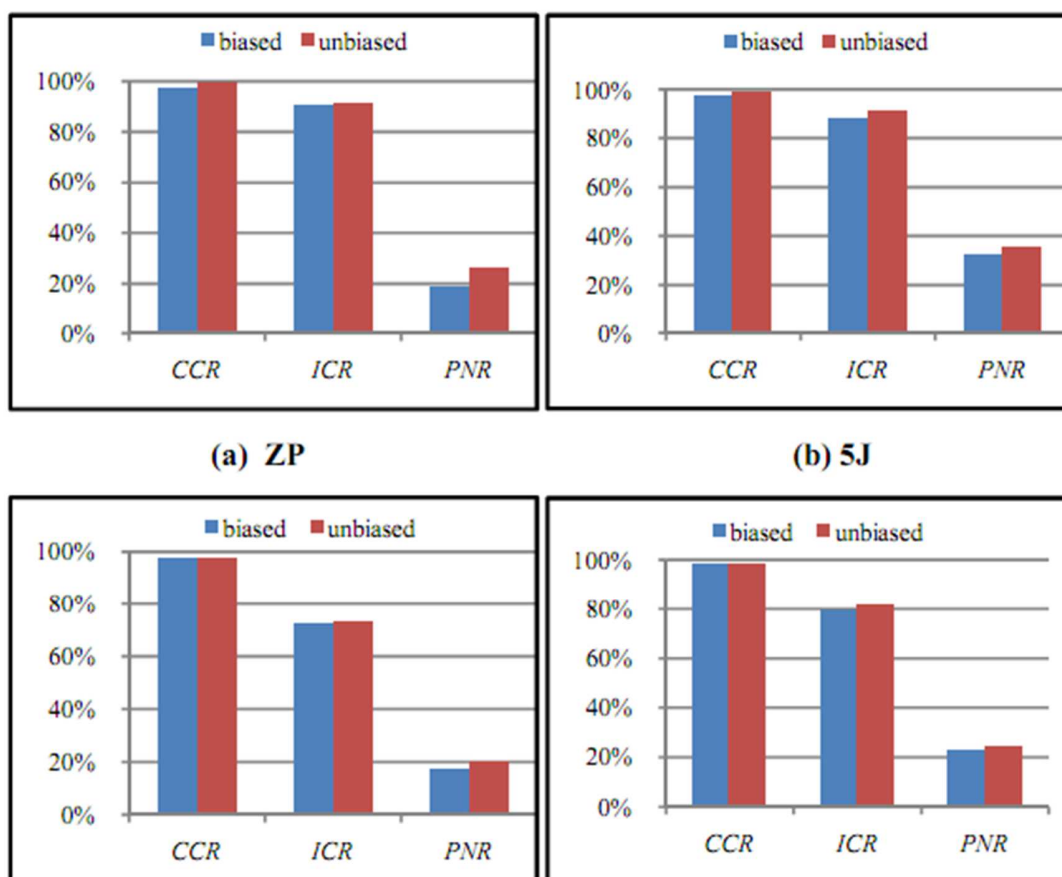
مطلب دلالت بر این دارد که روش مطرح شده برای نمونه های بدون تعصب عملکرد بالاتری دارد.

تحقیقات مرتبط

مورد آورده شده در قسمت 4 اولین تحقیق انجام شده در مورد تغییرات آرام در حوزه پایگاه اطلاعاتی وب است که برای استخراج اتوماتیک و آنالیز عناصر سطحی و تایید جستجو ها از طریق این سطوح جستجو ارائه شده اند. در قسمت 7 تحقیقات مربوط به ساختار جستجوهای کلیدی برای به دست آوردن مستندات از مجموعه پایگاه های اطلاعاتی وب آورده شده اند. در قسمت 8 نیز تکنیک های بررسی و انتخاب لغات کلیدی برای دانلود محتوای متن از پایگاه اطلاعاتی وب داده شده اند. در بخش 5 هم مطالبی در مورد کاهش مشکل انتخاب یک مجموعه جستجو از یک نمونه از منابع اطلاعاتی برای مسئله پوشش مجموعه بیان شده است. تحقیقات آورده شده در قسمت 6، تغییرات آرام در پایگاه اطلاعاتی وب را به صورت یک مجموعه پوششی مدلسازی کرده و یک مجموعه الگوریتم جدید برای بیان این مسئله ارائه می کند. تمام تحقیقات انجام شده در بالا متمرکز بر چگونگی ایجاد تغییر در کل پایگاه اطلاعاتی هستند که برای پایگاه های اطلاعاتی دینامیک کاربردی نمی باشد. یک روش متفاوت با نام مسئله افزایش تغییر را مطرح کرده ایم. اگر چه این روش در امتداد روشهای قبلی است، کاربرد آن در ترکیب اطلاعات واقعی غیر قابل اجتناب است. نمونه گیری از پایگاه اطلاعاتی وب همانند ایجاد تغییر در پایگاه اطلاعاتی است که شامل برداشت اطلاعات و جستجو از طریق ایجاد تغییر است و هدف نمونه گیری پایگاه اطلاعاتی وب هم به دست آوردن بخش کمی از گزارشات است که به صورت یکنواخت در پایگاه اطلاعاتی وب پخش شده اند. در قسمت 17 اولین تحفیف در مورد نمونه گیری اطلاعات وب داده شده است. در

این روش، تکذیب احتمالی بر اساس تکنیک های مورد استفاده برای ارتقای کیفیت و اندازه مورد استفاده قرار می گیرد و اندازه ها نمونه ها می توانند با تنظیم پارامترها تغییر کنند.

یک حوزه مهم و مرتبط دیگر ، استخراج اطلاعات وب است که یک مولفه کلیدی در ایجاد تغییرات آرام است. راه حل های بسیاری برای این مسئله مطرح شده است. بخش های 9 و 10 مسئله استخراج تمام اتوماتیک داده ها را با بررسی الگوهای موجود از نمونه های تولید شده بررسی کرده اند. بخش 11 ، از اطلاعات مربوط به گزارشات کامل صفحات گزارش شده که از نتایج حاصل از شناسایی و استخراج داده ها به دست آمده اند استفاده می کنند. شایان ذکر است که استخراج اطلاعات وب در راستای جستجو مسئله در این تحقیق مورد بررسی قرار گرفته است. علاقه زیادی به تحقیق در مورد درک معنای سطوح جستجو در پایگاه های اطلاعاتی ساختار بندی شده وجود دارد. قسمت 12 ، ترکیب کننده WISE را معرفی می کند که از اطلاعات جامع مثل عناوین داده ها و مقادیر پیش زمینه برای شناسایی اتوماتیک شاخص های متناسب استفاده می کند. در قسمت 13 ، یک نمودار تناسب مثال محور آورده شده است تا نمودار شاخص ها را ترسیم کند. در قسمت 14 ، از مدل های آماری برای یافتن نمودار پنهان از طریق آنالیز پیدایش اسم شاخص ها استفاده شده است.



VII. نتیجه گیری و تحقیقات آینده

دینامیک زیاد پایگاه اطلاعاتی باعث شده است تا روش ایجاد تغییر یک روش غیر کاربردی برای حفظ ثبات داده ها در بین پایگاه های اطلاعات محلی وب دیپ و ترکیب اطلاعات سیستم و پایگاه های اطلاعاتی شود. در این تحقیق ، ما مسئله تغییر در پایگاه اطلاعاتی رت بررسی کرده و یک روش موثر برای بیان این مسئله مطرح کردیم. تجربه گسترده در مورد پایگاه های اطلاعاتی جهان واقعی می تواند به صورت قابل توجهی هزینه تغییرات و کاهش پوشش را به حداقل برساند.

در تحقیقات آینده ، هزینه تغییرات را با ترکیب تعداد کلی مسائل در PNR ارتقا می دهیم که می تواند هزینه های مربوطه را به صورت مفهومی محاسبه کند.



این مقاله، از سری مقالات ترجمه شده رایگان سایت ترجمه فا میباشد که با فرمت PDF در اختیار شما عزیزان قرار گرفته است. در صورت تمایل میتوانید با کلیک بر روی دکمه های زیر از سایر مقالات نیز استفاده نمایید:

لیست مقالات ترجمه شده ✓

لیست مقالات ترجمه شده رایگان ✓

لیست جدیدترین مقالات انگلیسی ISI ✓

سایت ترجمه فا ؛ مرجع جدیدترین مقالات ترجمه شده از نشریات معتبر خارجی