

## تحلیل داده پایگاه ها:

### بررسی داده پایگاه های تهران بزرگ\*

نوشته: فاطمه فهیم نیا - نادر نقشینه  
مرکز اطلاعات و مدارک علمی ایران

#### چکیده

این مقاله به موضوعاتی می پردازد که ایجاد داده پایگاهی برای محدوده تهران بزرگ مستلزم آن هاست و به ویژگی های سودمند چنین امری نیز اشاره خاص دارد بعلاوه بر نیاز به استاندارد برای داده پایگاه ها تأکید می کند و نگاهی هم به تقویت کارآیی کارکردی این داده پایگاه ها دارد. این مقاله با ذکر تلاش هایی که در ایران برای شبکه کردن کتابخانه های ایران از طریق خدمات تحویل مدرک و داده پایگاه های مرکز اطلاعات و مدارک علمی ایران انجام گرفته به پایان می رسد.

در حال حاضر، صنعت رایانه بر همگی ابعاد زندگی ما اثر گذاشته و در حال تبدیل به یک صنعت بنیادین است. ولی باید به یاد داشت که سیستم های رایانه ای نیز حساس تر و پیچیده تر شده اند و خطاهای کوچک چه بسا که ضایعات سنگین و برگشت ناذیری رابه بار آورد. این سیستم ها با انعطاف پذیری بالای خود برای همین حد از دقت ساخته شده اند. این ویژگی استفاده کنندگان را به تقاضای مستمر برای تغییر در این سیستم ها واداشته است. از سوی دیگر این سیستم های نرم افزاری سرعت توسعه سیستم ها را که در مواردی به تغییر کلی در محیط عملیاتی منجر می شود افزایش می دهد و تقویت می کند. همگی این ها به نوبه خود باعث ظهور مسائلی در زمینه سازگاری و کارآیی می گردد. بدین ترتیب راهبری (۱) سیستم های نرم افزاری به یک مسئله ویژه بدل می شود. به موجب برخی ارقام، راهبری گاهی تا به ۶۷ درصد از هزینه اصلی نصب نرم افزار می رسد. در طول انجام این بررسی، دریافتیم که مستندات نرم افزاری هنوز یک مسئله است. هم.م مهندسان ایرانی نرم افزار، اصولاً کم ترین اهمیت را برای این جنبه از تولید نرم افزار قائل اند. حتی در قرار دادهایی که به شرکت های نرم افزاری داده می شود حداکثر یک عبارت ظاهری و تشریفاتی درباره مستندات نرم افزاری دیده می شود.

در بیش تر موارد، کسانی که قار است راهبری نرم افزاری را به عهده بگیرند باید در طول اجرای آن حضور داشته باشند. رفتار رسمی این است که اگر نرم افزار درست کارکرد، ناراحت نشوید! اما خنده دار این که هر اشکالی که درطراحی قطعات نرم افزاری وجود داشته باشد، به هزینه های بعدی راهبری تبدیل [ و به هزینه های دیگر خریدار، بار ] می شود. تعداد باز هم کم تری از شرکت ها

\* - Fahim

th FID Congress and conference, Jaipur & New delhi, 11-17 Ocgober

1998.

هستند که متعهد باشند به این که عمده هزینه های راهبری را [با دقت کردن] در ساخت محصول، به حداقل برسانند. برعکس، قاعده عمومی هر چه کم تر کردن زمان اجرا است، نه زمان راهبری. بنابراین گزارش تحقیقی که از سوی ISIS انجام شده در برخی موارد حتی انتخاب نادرست زبان برنامه نویسی برای کدهای اصلی ممکن است باعث افزایش سرسام آور هزینه های راهبری شود اما مهم ترین موضوع شاید بی توجهی تولید کننده نرم افزار به این نکته بدیهی باشد که سیستم های نرم افزاری قرار نیست تا ابد عمر کنند. در درون مراکز اطلاع رسانی ایرانی، تعداد اندکی هستند که استانداردهای لازم الاجرا و فرایندهای [کنترل و] تأیید را در عرصه سیستم های نرم افزاری به کار بندند. با افزایش خودکار سازی در مراکز اطلاع رسانی ایرانی و با وجود گرایش کنونی در ساخت داده پایگاه های داخلی و در دسترس گذاشتن آن ها حتی برای استفاده کنندگان راه دور، تحقیقی در موضوع پذیرش استانداردهای نرم افزار داده پایگاه ها انجام شد. شهرداری تهران در پی آن بود که امکان اداره یک داده پایگاه بسیار بزرگ شامل ۱۵۰ واحد اطلاعاتی یا داده پایگاه فرعی را بررسی کند. این داده پایگاه ها را نزدیک به چهل مرجع جداگانه در دست ایجاد و خدمات دهی داشتند و عرصه هایی همچون حوادث، رویدادهای فرهنگی، مراکز پزشکی، فهرست قیمت ها، کتابخانه ها، و حتی خدمات سفارتخانه های خارجی در این داده پایگاه ها ملحوظ شده بود. اساساً هدف شبکه محدود تهران بزرگ تأمین دسترسی به این اطلاعات به دو شکل زیر است:

الف) با واسطه: از طریق کیوسک های اطلاعات؛

ب) بیواسطه: از طریق شبکه رایانه ای؛ مثل نمونه ای که "مؤسسه ندا رایانه" راه اندازی کرده و پیشتر در گسترش دسترسی همگانی به اینترنت در ایران است. صرفه جویی بالقوه ناشی از کاهش تعداد سفرهای درون شهری که با کاهش آلودگی و نیز بهره گیری بهتر از منابع ارتباطی همراه است به شبکه محدود تهران بزرگ جاذبه و مقبولیت بخشیده است. وقتی کسی از ایجاد سرویسی برای ارائه خدمات به حدود ۱۰۰۰۰۰ نفر در روز صحبت می کند، نقش استانداردها بسیار برجسته تر می شود. هماهنگ کنندگان داده پایگاه نیازمند بررسی و به هم پیوستن یک برچسب مجازی به دست آمده از سیستم های داده پایگاهی هستند. در بیش تر موارد مبدل های واسطه نرم افزاری خاصی باید نوشته می شد تا بتوان عناصر داده ای را از داده پایگاه های مختلف صادر یا به آن ها وارد کرد. هیچ چیزی بیش از بخش کتابخانه ای نمی تواند وسعت کار را به روشنی تصویر کند. عمده کتابخانه های تهران در زمان انجام این بررسی روی هم رفته از چهار سیستم متفاوت خودکار سازی کتابخانه استفاده می کردند که همدیگر را نمی خواندند. ظاهراً در آن زمان هنوز اخبار Z ۳۹/۵۰ به صحنه کتابداری ایران نرسیده بود. بنابراین، این بررسی اساساً در دو مسیر انجام شد. در مسیر اول کوشش بر این بود که میزان پایبندی داده پایگاه های جاری به مجموعه استاندارد های موجود، از قبیل استانداردهای ایزو و یونی سیست را مشخص کند. این کار میزانی برای اندازه گیری دامنه و عمق اصلاحات نرم افزاری لازم برای ادغام داده پایگاه ها به دست می دهد. مسیر دوم در پی رسیدن به استانداردهای جدیدی برای تضمین سازگاری و عملیات پذیری بین سیستمی بود. همه

این بخش ها در معرض الگوی کسفیت بوئن (۲) قرار داشتند که بر معیار های نرم افزاری مشخص و تعریف شده مبتنی است. این معیار ها عبارت بودند از:

- صحت؛
- اطمینان پذیری؛
- بازدهی؛
- انسجام؛
- کارآیی؛
- راهبری پذیری؛
- نرمش پذیری؛
- آزمایش پذیری؛
- حمل و نقل پذیری؛
- قابلیت استفاده مجدد؛
- قابلیت عملیات متقابل؛

“وراپایگای” (۳) موضوع مهمی در بررسی این بخش های اطلاعاتی بود. مثلاً داده پایگاه هایی را که به وسیله مراجع مالی همچون بانک ها و مؤسسات پس انداز و قرض الحسنه ایجاد می شد همگان نمی توانستند مشاهده و بررسی کنند. با این حال، محافظ های مناسبی که به منظور تأمین محرمانگی تدارک (۴) شده اند قابل دسترسی اند. برای خدمات کنترل اعتبار [مالی]، ابزار تأیید نیرومندی تهیه شده که با اطلاعات مالی همراه می شود و از طریق “اداره ثبت شرکت های ایرانی” قابل دسترسی است. در عین حال، فرایند بررسی علاوه بر تحلیل تأثیر تک تک داده پایگاه ها، نیازمند ارزیابی تأثیر آنها در ترکیب با دیگر داده پایگاه ها نیز هست. این کار می تواند چشم انداز جذابی را نتیجه دهد که توسعه سریع داده پایگاه های سفارشی با میزان بالاتری از قابلیت اجرایی بلاواسطه را تضمین می کند. مثلاً ادغام “داده پایگاه شرکت نامه رسانی پیک بادپا” و “داده پایگاه کد پستی شرکت پست ایران” داده پایگاهی را نتیجه می دهد که می توان از آن برای تحویل کارآمد و صرفه مندانه کالاهای حساس در تهران بهره گرفت. عمده داده پایگاه های دایر در محدوده تهران بزرگ از سوی دولت ایران حمایت مالی می شوند. اما فقدان کامل استانداردسازی در ایجاد آن ها به وضوح به چشم می خورد. معدودی از این داده پایگاه ها اعلام کرده اند که مجموعه استاندارد های گروه ایزو ۹۰۰۰ را رعایت می کنند. با آن که ISIRI ۲۹۰۰ به “اصول استاندارد ایرانی برای تبادل اطلاعات” (ISCLL) می پردازد، برخی طراحان داده پایگاه ها بوجه بایسته ای به آن نکرده اند و در نتیجه، وقتی این داده پایگاه ها قرار بود به کاربردهای مبتنی بر وب تغییر شکل دهند، اشکالات جدی بروز می کرد. درباره میزان سازگاری با رفتار اطلاع جویی استفاده کنندگان هیچ استاندارد موجودی در دست تهیه نبود. این امر زماین که شخص بخواهد داده پایگاهی را برای گروه بزرگ تری از مخاطبان - با پس زمینه های متفاوت قومی- فرهنگی و تحصیلی، نظیر

آنچه در میان شهروندان تهرانی می بینیم- ارتقای سطح دهد، به مسئله ای جدی بدل می شود. در برخی از جنبه های ناوبری اطلاعات و بازنمایی تصویری معناساختی به این امر پرداخته شده داده پایگاه های بسیار معدودی از ویژگی های کمی استاندارد یا کمک های تصویری تقویت شده ناوبری بهره می گیرند "ماورا" که یک عملگر شبکه دیگر در ایران است، در حال حاضر تحقیقی را درباره ایجاد کمک های دقیق نرم افزار مبنا دنبال می کند. اما سال ها طول می کشد تا این تلاش ها به نتیجه ملموسی برسد.

### یافته های این بررسی

در حدود ۷۳/۶ درصد داده پایگاه های محدوده تهران بزرگ برای عملیات با دستگاه های AT۲۸۶ یا دستگاه های قوی تر، و ۱۴/۵ درصد برای دستگاه های AT۳۸۶ یا بالاتر طراحی شده اند. ظاهراً طراحان می خواسته اند از سخت افزارهای خاص و پردازشگرهای قدرتمند اجتناب کنند. در حدود ۵۲/۷ درصد از داده پایگاه های پژوهشی از ترکیبی از بسته های حاضر و آماده، و نرم افزارهای سفارشی استفاده کرده اند و ۳۶/۶ درصد از نرم افزار یونسکو، یعنی SDC/ISIS به عنوان نرم افزار اصلی بهره برده اند. ۶۹/۲ درصد این داده پایگاه ها دارای پوشش پنج ساله اند و نزدیک به ۷۰ درصد آن ها منضمماً روز آمد می شوند. داده پایگاه های چند رشته ای ۴۲ درصد داده پایگاه های موجود در محدوده تهران بزرگ را تشکیل می دهند. در مرتبه بعد رشته مهندسی (۲۰ درصد)، علوم انسانی (۱۴/۶ درصد)، پزشکی (۳/۶ درصد)، و علوم پایه (۱۸ درصد) قرار دارند. ۷۵ درصد داده پایگاه ها کتابشناختی و تنها هفت درصد آن ها متن کامل هستند. ۶۴ درصد داده پایگاه های موجود در تهران بزرگ دارای پوشش کشوری اند و اطلاعات خود را از تمام بخش های کشور گردآوری می کنند. بدین ترتیب گمان می رود که تهران دارای بیشترین میزان تمرکز از حیث نهاد های اطلاع رسانی دولتی باشد. گزارش کامل این بررسی در سال ۱۹۹۶ ارائه شد. در سال ۱۹۹۷ مشاوران پژوهشی در ندارایانه، ماورا، و تعدادی از نهادهای اطلاع رسانی خصوصی در ایران تصمیم گرفتند بستری برای آزمودن راه حل های احتمالی برای این مسئله ایجاد کنند. قرار بر این شد که این پایگاه ها هر کدام یک محصول به شمار آید و بر همین قیاس نیز با آنها رفتار شود. تصمیم گرفته شد که پروژه های غیر فعال OCLC ندا رایانه و "فرهنگ ۱" ماورا از فعال شوند و به گروه منتخبی از داده پایگاه ها متصل گردند. در زمان نوشتن این مقاله، پروژه OCLC ندارایانه موفق شده است که با ۱۰ کتابخانه همکار سرویس قابل قبولی را در شبکه خود ارائه کند. این سرویس، داده پایگاه های کتابخانه ملی ایران و مرکز اطلاعات و مدارک علمی ایران، و خدمات تحویل مدرک این مرکز را نیز شامل می شود. در هر حال، کار ایجاد یک میانجی نرم افزاری هوشمند هنوز در مرحله اثبات است.

### پی نوشت ها

1. Maintenance
2. Boenm quality model

3. Meta- basing
4. Privacy