

بخش اول: ایجاد شاخص وارون لغت-سند

در این بخش وقوع لغات در مقالات مختلف تحت بردار هایی باید ذخیره می شدند.

حجم بالای این Posting ها مانع نگهداری کامل آن ها بر روی حافظه ی اصلی بود. از طرفی به دلیل تعداد بالای لغات دسترسی به فایل posting برای هر بار به روز رسانی آنها هم بسیار وقت گیر بود.

راهکاری که برای این موضوع استفاده شد LRUcache با گنجایش معقولی از حافظه ی اصلی قابل استفاده بود و به این صورت با نگاه داشتن لغات پرکاربرد در حافظه سرعت پردازش به طور خوبی بهبود پیدا کرد.

برای دسترسی مستقیم به posting ها تحت هر کدام در یک فایل مجزا نگه داری می شوند.

تودو!

بخش دوم: به دست آوردن وزن کلمات

این بخش به دلیل عدم امکان محاسبات حین پردازش اولیه و نیاز به بازگزاری مجدد مقالات زمان گیر بود. با استفاده از ساختار داده های مناسب، pre-calculate کردن بعضی پارامتر ها و log باینری سریع سرعت و میزان حافظه ی مصرفی این بخش بهبود یافت.

وزن کلمات بعد از محاسبه به شاخصی که از فاز دوم تولید شده بود افزوده می شوند.

یک راهکار برای کاهش حجم این وزن ها تبدیل آنها به مقادیری نسبی بود که به دلیل عدم اطمینان از صحت این روش در پیاده سازی اعمال نشد.

بخش سوم: خوشه بندی اسناد

شاید یکی از چالش برانگیز ترین بخش ها مربوط به خوشه بندی اسناد بود ! برای پیاده سازی از خوشه بندی سلسله مراتبی استفاده شد.

معضل پیش آمده در این بخش عدم امکان نگاه داشتن تمامی خوشه ها در حافظه اصلی بود. راهکار استفاده شده در این بخش رهاکردن سریع زیر شاخه ها بعد از پرشدن شاخه بود.

بخش چهارم و پنجم : وارد کردن درخواست کاربر و بازگرداندن نتیجه

در این بخش مجدداً خوشه های اصلی در حافظه بازگزاری میشوند و درخواست کاربر بعد از تبدیل به داده های مناسب و وزن دادن به لغات با خوشه های اصلی و سپس زیر شاخه ها و نهایتاً اسناد مورد نظر در ۳ لایه مقایسه شده و ۱۰ نتیجه ی برتر در رابط تحت وب نمایش داده می شود.

ساختار داده ی مورد نظر برای این لایه ها TreeSet بود چون بهترین زمان مرتب سازی بر اساس میزان تطبیق را در اختیار قرار می داد.

در زیر نمونه ای از محیط کاربری قابل مشاهده است :

صفحه ی اصلی

موتور جست و جوی پویا

سلام ایران

جست و جو

جست و جو شده در ۱۱۸۴ هزارم ثانیه

مقاله شماره ی ۲۲۹

... رزندی برومند برای خراسان و خدمتگزاری دانشمند برای ایران تربیت کرد احمدی نه فقط عالمی ادیب و د

مقاله شماره ی ۱۷۲

... اسبت برپایی ششمین نمایشگاه دوسالانه طراحان گرافیک ایران فراتر از یک دهه است که به غیر از کارش

مقاله شماره ی ۳۱۲

... ملل روزپنشنیه برگزار شد داریوش مهرجویی کارگردان ایرانی نیز شرکت داشت کارگردان های دیگری از

مقاله شماره ی ۷۰

... زد مازیار شیبانی فر بر نامه های ماه فرهنگی و دوستی ایران و ژاپن در تهران در ماه فرهنگی و دوستی

مقاله شماره ی ۴۳

... و موسیقی در اجتماع سهم دارند اداری کردن موسیقی در ایران که از حدود شصت سال قبل شروع شد تا امر

مقاله شماره ی ۹۶

... ای دیگر به دلایلی که برنگارنده چندان روشن نیست در ایران معرفی نشده است شاید زمان آن رسیده باش

اگر در جامعه ای امن نباشد دستیابی به شعار استقلال آزادی و جمهوری اسلامی با مشکل مواجه می شود شهیدی با اشاره به توفیق مطبوعات در سال جاری خاطرنشان کرد این مساله اتفاقی مهم و پیچیده در کشور بود و باید هیاتی متشکل از قوه قضا به مطبوعات و جامعه شناسان این مقطع زمانی را بررسی کرده و نظرات خود را به مردم انعکاس دهند تامین اجتماعی از لایحه بازنشستگی پیش از سخت حمایت می کند موعد مشاغل گروه اجتماعی سازمان تامین اجتماعی از لایحه دولت در خصوص بازنشستگی پیش از موعد مشاغل سخت و زیان آور حمایت جدی می کند طاهر موهبتی معاونت حقوقی و امور مجلس سازمان تامین اجتماعی دیروز در جمع خبرنگاران با اعلام این خبر گفت سازمان تامین اجتماعی در تلاش است تا این لایحه هر چه سریعتر در دستور کار مجلس شورای اسلامی قرار گیرد وی افزود کارفرمایان کلیه کارگاههای مشمول قانون تامین اجتماعی به موجب این لایحه مکلفند ظرف دو سال از تاریخ تصویب این قانون جهت ایمن سازی عوامل و شرایط محیط کار مطابق حد مجاز و استانداردهای مشخص شده در قانون کار و آیین نامه های مربوطه اقدام نمایند موهبتی اضافه کرد از تاریخ تصویب این قانون چهاردرصد به حق بیمه کارفرمایان افزوده خواهد شد وی همچنین گفت بنابراین لایحه تشخیص مشاغل سخت و زیان آور به موجب آیین نامه خواهد بود که وزارتخانه ها و سازمانهای ذیربط تهیه و به تصویب هیات وزیران خواهند رساند وی در بخشی دیگر از این مصاحبه به بودجه سازمان تامین اجتماعی در سال پرداخت و گفت دولت در سال بخشی از بدهی خود به سازمان تامین اجتماعی را پرداخت می کند وی گفت در واقع دولت صد میلیارد تومان از بدهی های خود را از طریق واگذاری سهام متعلق به خود در شرکتهای دولتی و غیردولتی پرداخت خواهد کرد وی حجم بدهی های دولت را نزدیک به میلیارد تومان ذکر کرد و گفت در بودجه سال به سازمان تامین اجتماعی اجازه داده می شود که در سال جهت ارائه تسهیلات ارزان قیمت ساخت و خرید مسکن ودیعه مسکن ازدواج دانشجویی و هزینه مربوط به خدمات مشاوره ای در زمینه بهداشت روانی فردی اجتماعی میلیارد ریال از محل درآمد و ذخایر خود با تصویب شورایعالی سازمان تامین اجتماعی به کارگران و مستمری بگیران شرکتهای تعاونی مسکن کارگران از طریق بانک رفاه کارگران پرداخت نماید مدیرکل مطبوعات داخلی وزارت فرهنگ و ارشاد اسلامی به ازای هر نفر در