

برگه ارزیابی

با سلام خدمت کارجوی عزیز؛

اکنون که در حال مشاهده این برگه هستید به معنای آن است که مرحله اولیه مصاحبه را با موفقیت گذرانده‌اید، فعالیت‌های زیر به منظور ارزیابی دقیق‌تر توانمندی و مهارت‌های شما در نظر گرفته شده و نتایج آن برای ادامه فرآیند مصاحبه ضروری بوده و در ارزیابی تان نقش به‌سزایی خواهد داشت؛ لذا شایسته است تا با تلاش زیاد، دقت عمل و کیفیت پاسخگویی بالا اقدام به تکمیل برگه ارزیابی نمایید. در انتها پس از انجام کارخواسته‌های زیر، مستندات پاسخ خود را در قالب فایل zip جمع‌بندی و ارسال نمایید.

«مسئله شماره ۱»

- ✓ به دنبال جمع‌آوری داده از یکی از بسترهای وب می‌باشیم، مسیر زیر را در بستر انتخابی خود طی نموده و گام به گام اقدام به مستندسازی و ارائه گزارش نمایید.
- ✓ گام‌های زیر امتیاز تجمیعی داشته و انجام گام‌های بیشتر امتیاز بیشتری برای شما در پی دارد.
- ✓ در صورت استفاده از حساب‌های کاربری رایگان، اشاره به مشخصات ورود در مستندات ضروری است.

گام	مهارت	فعالیت	توضیحات / راهنمایی
۱	معماری داده‌ها	طرح کلان و نقشه ذهنی خود را برای پیاده‌سازی این پروژه ترسیم و گزارش نمایید.	نقشه می‌بایست با ابزارهای استاندارد مانند Microsoft Visio طراحی بشود.
۲	جمع‌آوری داده‌ها	اقدام به طراحی و برنامه‌نویسی خزنگری برای جمع‌آوری بستر انتخابی بنمایید. موجودیت‌های بستر انتخابی را مشخص کرده و برای هر کدام برنامه خزشی مشخص نمایید (مانند: پست، تایم‌لاین، تیتیر، دنبال‌کننده و دنبال‌شونده، لایو، استوری، دوستان و ...).	بسترهای پیشنهادی برای جمع‌آوری (به ترتیب ارزش امتیازی): اینستاگرام، لینکدین، فیسبوک، توییتر، تلگرام، تارنما، وبلاگ. بستر مورد نظر را به هر دو روش API و Scrap ارزیابی نمایید.
۳	جمع‌آوری داده‌ها	برنامه اجرایی خود برای مقیاس‌پذیری را شرح دهید.	بهره‌گیری از روش‌های موازی‌سازی و کانتینرها مزیت محسوب می‌شود.
۴	کار تیمی	مراحل اجرایی خود را در یک بستر کنترل پروژه ثبت نمایید و گزارش کار و پیشرفت خود را در آن لحاظ نمایید.	با استفاده از حساب کاربری رایگان سامانه مدیریت پروژه و تیم پیگیر (peygir.com) یا

			جیرا (Jira) این فعالیت را انجام دهید.
۵	مستندسازی	برنامه‌نویسی خود را در یک بستر کنترل نسخه ثبت و پیش ببرید.	با استفاده از ویژگی Git حساب کاربری رایگان سامانه مدیریت پروژه و تیم پیگیر (peygir.com) یا گیت‌هاب (Github.com) این فعالیت را انجام دهید.
۶	ترابری داده‌ها	با استفاده از یک بستر انتقال پیام (Message Broker) داده‌های جمع‌آوری شده توسط خزسگر خود را دریافت نمایید.	بنا به نیاز یکی از ابزارهای ActiveMQ, RabbitMQ, Kafka یا Spark Streaming انتخاب کنید.
۷	ترابری داده‌ها	با استفاده از یک میان‌افزار اقدام به دریافت پیام‌ها نموده و به صورت در لحظه عملیات پردازشی را بر روی آنان انجام دهید و آن را آماده ذخیره‌سازی در پایگاه داده نمایید.	با استفاده از ابزارهایی مانند Apache Nifi یا Airflow این کار را انجام دهید و برای عملیات پردازشی هم از وب‌سرویس‌های پردازش متن زبان فارسی رایج مانند واکاو یک در apieco.ir استفاده کنید.
۸	انبارش داده‌ها	با توجه به نوع داده جمع‌آوری شده mapping / schema لازم را طراحی و بیان نمایید و سپس اقدام به درج داده‌ها در پایگاه داده نمایید.	ذخیره‌سازی متون به شکل جستجو محور (در پایگاه داده Elasticsearch)، موجودیت‌ها به شکل گراف و داده خام در پایگاه داده ستونی مزیت محسوب می‌شود. در صورت عدم امکان موارد بالا، نتایج در SQLite یا سایر RDBMSها ذخیره شود.
۹	پایش داده‌ها	فرآیند جمع‌آوری خود را به صورت نظام‌مند مورد پایش (monitoring) قرار دهید و نتیجه را گزارش دهید.	از ابزارهای BI یا APM مانند Logstash و Kibana می‌توانید استفاده کنید.
۱۰	کلان داده‌ها	معماری فرآیند طی شده را با استفاده از ابزارهای کلان داده بازطراحی و گزارش نمایید.	بهره‌گیری از چارچوب Hadoop و ابزارهای ارائه‌دهنده آن مانند Cloudera پیشنهاد می‌شود. معرفی سایر ابزارها و روش‌ها مزیت محسوب می‌شود.