



دسته‌بندی متون پروژه مبانی برنامه‌سازی

زمستان ۹۵

فاز ۱

در این فاز شما با دریافت فایل‌های RSS از سرور پروژه آن‌ها را پردازش، ساختاربندی و ذخیره می‌کنید.

زمان‌بندی

مهلت انجام این فاز از ۴ تا ۱۸ دی است.

ساختار فایل‌ها

شما باید عملیات پردازش روی فایل‌های URL دریافت شده را در یک فایل و عملیات مربوط به دریافت اطلاعات از سایت، ذخیره‌سازی آن در فایل و فرستادن اطلاعات را در یک فایل دیگر انجام دهید.

برای استفاده از اطلاعات یک فایل در فایل دیگر باید آنرا include کنید؛ مثلاً اگر فایل‌های a.c و b.c در یک آدرس هستند و میخواهید از توابع فایل a.c در فایل b.c استفاده کنید به این شکل عمل کنید

```
#include "a.c"
```

RSS چیست؟

RSS قالبی برای انتشار محتواهای اینترنتی است که مدام در حال به‌روزرسانی هستند. سایت‌های خبری و وبلاگ‌ها از این قالب برای انتشار محتوای خود و در دسترس قرار دادن آن استفاده می‌کنند. یک فایل RSS شامل اطلاعات و فراداده‌های مختلف مانند تاریخ انتشار، نام نویسنده و تیتراست.

RSS به منتشر کنندگان اطلاعات، مانند سایت‌های خبری، این امکان را می‌دهد که داده‌هایشان را به صورت خودکار در قالب فایلی XML که با سیستم‌های مختلف سازگار است منتشر کنند. RSS همچنین به کاربران این امکان را می‌دهد تا به صورت منظم اطلاعات مورد نظرشان را از سایت‌های مختلف دریافت کنند.

دریافت اطلاعات

در این فاز شما ابتدا تعدادی URL صفحات RSS را با فرستادن یک درخواست GET به آدرس

<http://www.fop-project.ir/news/get-urls>

با شناسه‌ی نام کاربری و رمز عبور تیمتان دریافت کنید. این URL ها با فاصله از هم جدا شده اند. سپس به آدرس هر کدام از URL ها یک درخواست دیگر باید بفرستید تا فایل RSS مربوطه را دریافت کنید. هر فایل شامل تعدادی خبر است که در قالب RSS نوشته شده اند. توصیه می‌شود که در این مرحله فایل‌های RSS ای را که دریافت می‌کنید در سیستم خود ذخیره کنید تا هر بار هنگام اجرا شدن برنامه مجبور به دریافت فایل‌ها از سرور نباشید. همچنین چون از این فایل‌ها در فازهای بعدی هم باید استفاده کنید ذخیره کردن آنها منطقی است! می‌توانید از امکانات زبان C برای کار با فایل استفاده کنید.

پردازش اطلاعات

هر فایل RSS ای که شما دریافت می‌کنید دارای ساختار مشخصی است. هر خبر با یک تگ item مشخص شده‌است و سایر اطلاعات در قالب تگ‌های دیگر در این تگ مشخص شده‌اند. داده‌هایی که شما باید از هر خبر استخراج کنید در قالب تگ‌های صفحه‌ی بعد آمده‌اند:

- تگ title که دربردارنده تیتر خبر است.
- تگ link که شامل پیوندی به صفحه خبر است.
- تگ pubDate که تاریخ انتشار خبر را مشخص می‌کند.
- تگ description که حاوی خلاصه‌ای از خبر است.
- و در نهایت تگ‌های category که دسته‌بندی‌های مربوط به خبر را مشخص می‌کنند.

دقت کنید که اطلاعات ذخیره شده بین تگ‌ها لزوماً به آن فرمتی که شما می‌خواهید نیست و ممکن است خود شامل اطلاعات اضافه دیگری باشد! مثلاً ممکن است در تگ description آدرس یک عکس در یک تگ آمده باشد که قرار آن در متن description ای که برای سایت می‌فرستید منطقی نیست! برای همین حتماً چند نمونه از فایل‌های RSS ای که فرستاده‌شده را به دقت بررسی کنید تا هنگام پردازش آن‌ها با مشکلی مواجه نشوید! یا مثلاً محتویات تگ‌های کتگوری فرمت خاص خود را دارند.

نکته مهم: دقت کنید که بعضی از link ها با پیشوند http و بعضی با https آمده‌اند. شما باید همه‌ی لینک‌ها را پیشوند https ذخیره کنید تا در فازهای بعدی به مشکل نخورید. شما باید این داده‌ها را برای هر خبر استخراج کنید و هر خبر را در قالب یک درخواست POST به همان آدرسی که URL ها را گرفتید بفرستید. دقت کنید که خبرها را باید با فرمت JSON کد کنید و بفرستید.

ذخیره‌سازی

بعد از پردازش اطلاعات فایل rss می‌توانید خبرها را در struct ذخیره کنید. دستوری در C وجود دارد که محتوای یک struct را به طور کامل داخل یک فایل قرار می‌دهد و از همان فایل می‌خواند. این دستور را پیدا و از آن استفاده کنید. البته طبیعتاً چنین فایلی خیلی انسان فهم نیست ولی همین که C آنرا می‌فهمد کافی است!

فرستادن اطلاعات

جزئیات قسمت data درخواست شما به این شرح است: (البته توصیه می‌کنیم با جستجو در اینترنت در مورد فرمت JSON که در تبادل اطلاعات بسیار پرکاربرد است تحقیق کنید) در فاز قبل شما داده‌های POST خود را به صورت عادی فرستادید key=value (که به این قالب form-data می‌گویند و در فرم‌های اینترنتی پر کاربرد است مثل فرم لاگین همین سایت خودمان که داده‌های شما را مثلاً به شکل username=team1&password=1234314 می‌فرستد) در این فاز به دلیل امکانات بسیار بهتر قالب JSON و انسان فهم بودن بیشتر آن خبرهای خود را در این قالب می‌فرستید:

```
{
  "title": "I met my dad",
  "date": "27 Dec 2014",
  "description": "I met my dad that day",
  "categorized": true,
  "categories": ["Family", "Politics", "Father"]
}
```

قالب بسیار ساده است و همانطور که می‌بینید ساختارش مثل form-data مجموعه‌ای از key و value هاست با این تفاوت که چند نوع متغیر در این قالب تعریف شده است: عدد، رشته، آرایه، boolean و object. تمامی کلیدها (مثل categories, title, description) به صورت رشته‌های ثابت اند و حتماً باید داخل “ باشند. در مثال بالا مقادیر متغیرهای متناظر title و date و description از نوع رشته اند در حالی که مقدار متناظر

categorized از نوع boolean است (به این دقت کنید که عبارت true بر خلاف قبلی‌ها داخل “ قرار ندارد!). در آخر مقادیر متناظر categories همانطور که از اسمش پیداست از نوع آرایه است. آرایه ای از رشته‌ها. کل موجودیت خبر هم متغیری از نوع Object است. اصولاً برای آرایه‌ها از گروه و برای object ها از آکولاد استفاده می‌شود. هم‌چنین دقت کنید که درخواست‌های POST شما باید از نوع ajax باشد. برای این کار در Header درخواست POST تان عبارت X-Requested-With: XMLHttpRequest را قرار دهید.

چند نکته

- کل عبارت بالا یک رشته است. یعنی در قسمت data ریکوئست خود باید یک رشته ی به فرمت جیسون یا json string بفرستید. این رشته در میان خود کاراکتر “ دارد ولی در مجموع یک رشته است.
- در هنگام ارسال درخواست باید به نوعی مشخص کنید که داده های خود را مثل دفعه قبل به صورت form-data نمیفرستید و از قالب جیسون استفاده می کنید. به این موضوع content-type می گویند و پس شما باید در قسمت Header درخواست خود علاوه بر Authoriation, Content-Type را هم مشخص کرده و آن را برابر با application/json قرار دهید.
- مقادیر عنوان، تاریخ و توصیفات و دسته بندی ها را بر حسب خبر مدنظر خود پر کنید ولی دقت کنید که کلیدها را عینا درست وارد کرده باشید (مثلا در خود فایل rss ممکن است تاریخ با تگ pubDate مشخص شده باشد ولی شما در هنگام فرستادن درخواست باید تاریخ را با کلید date بنویسید). همچنین مقدار categorized را همواره برابر true قرار دهید (به این معنی است که خبرهایی که پردازش کردید دارای تگ های category بوده اند).
- یک پیشنهاد منطقی این است که قبل از فرستادن خبر یک بار رشته json تولید شده را ببینید و بررسی کنید.
- تعداد دفعاتی که می توانید به سایت درخواست بدهید محدود است (در دقیقه). پس اگر درخواست های زیادی بدهید که رشته ی json شما درست نباشد ممکن است برای دقایقی سایت IP شما را بلاک کند!

مشاهده ی نتیجه تلاش ها در سایت

در آخر در صورتی که در response سرور خطایی نداشتید، یعنی خبرها با موفقیت ذخیره شده و شما می توانید با ورود به سایت و انتخاب قسمت search از نوار بالای صفحه، دسته بندی های خبرهایتان را وارد کرده و اخبار را مشاهده کنید. البته باید حتما عین کلمه را جستجو کنید و طبعاً نمی توان انتظار داشت موتور جستجوی ما امکاناتی مثل تصحیح خطا و یا پیشنهاد کلمه داشته باشد! پس اگر به طور دقیق خود کلمه را (با رعایت حروف کوچک و بزرگ و فاصله در آخر کلمه) جستجو نکنید جواب خاصی نخواهید گرفت!

اما اگر کلمه را درست وارد کرده باشید فهرست تمامی خبرهایی که شما فرستادید و در دسته بندی مورد نظر هستید به همراه سایر اطلاعاتشان مثل تاریخ و توضیحات برایتان به نمایش درخواهد آمد. دقت کنید که در هنگام جستجو باید حتما تیک مربوط به categorized را در پایین searchbox موجود بزنید!

دقت کنید که بیشینه تعداد خبرهایی که هر تیم می تواند در سایت داشته باشد محدود است و اگر چندین بار همه خبرها را بفرستید تعدادی از خبرهایی که قبلاً فرستادید حذف می شوند تا تعداد خبرهایتان از سقف مجاز بیشتر نشود.

آپلود در کوئرا

در آخر روی آخرین کامیتان یک tag بگذارید و فایلهایتان را طبق ساختار گفته شده در کوئرا آپلود کنید. مانند فاز قبل باید کد شما در هنگام تحویل حضوری تمام مراحل را مجدداً اجرا کند. طبعاً در هنگام تحویل حضوری نمیتوانید دست به کد بزنید پس حتماً مراقب این موضوع باشید که کد کاملی را به عنوان آخرین کامیت فاز ۱ داشته باشید.

سوالاتان را هم در کوئرا بپرسید!