بسم الله الرحمن الرحیم

راهنمای developer ماژول web kit

آذر 97

**فصل 1**

# **1-1 نگارش**

در جدول زیر تاریخچه نگارش های مختلف ماژول آورده شده است.

|  |  |  |
| --- | --- | --- |
| تاریخ | نسخه | توضیحات |
| آذر 97 | 2.0 | نسخه اولیه document آماده شد. |

# **1-2- توضیحات اولیه**

ماژول web kit برای پیاده‌سازی ماژول‌های مبتنی browser فراهم شده است. این ماژول زبان Python و بر اساس ابزارهای موجود در binding کتابخانه QT (pyQt)، با کمک Java Script نوشته شده است و برای استفاده از آن یک wrapper به زبان python نوشته شده است که شامل متدهایی برای استفاده از web kit می‌باشد.

ورژن قبلی web kit نیز سازوکار مشابهی داشت با این تفاوت که web kit به زبان C++ توسعه داده شده بود و مستقیما از Qt استفاده می‌شد.

دلیل توسعه نسخه جدید، چالش‌های موجود در آن و امکانات حال حاضر web kit در ادامه آمده است.

# **1-3- دلیل پیاده سازی نسخه جدید**

1 - قدیمی شدن نسخه مرورگر استفاده شده در پیاده سازی قبلی

2 - پشتیبانی نکردن از برخی سایت‌ها و همچنین باز نکردن سایت‌هایی که از captcha جدید گوگل استفاده می‌کردند مانند linkedin

3 - امکانات بسیار زیاد و کاربردی نسخه جدید

4 - عدم وجود نمونه پیاده سازی شده با نسخه جدید webkit chromium

# **1-4- مشکلات پیش آمده در زمان شروع پیاده سازی**

1- نبود نمونه برای گرفتن ایده و پیاده‌سازی

2- مستندات برای زبان پایتون وجود نداشت و فقط مستندات C++ موجود بود، در حالت کلی مشکلی نداشت اما بعضی معادل‌سازی‌ها به دلیل تعاریف متفاوت دو زبان از شی گرایی و کار با objectها نیاز به نرمال‌سازی و variation & type converting داشت.

3- برخی متدها مانند اجرای کدهای جاوا اسکریپت که بخش اصلی Hrobot را تشکیل می‌دهند مانند قبل همزمان نبودند و به صورت غیر همزمان طراحی شده بود که با طراحی EventLoop و ارسال signal حل شد.

4- ارتباط بین پایتون و جاوا اسکریپت به صورت string برقرار بود و امکان دسترسی به object های یکدیگر را نداشتند که با پیاده‌سازی یک رابط برقرار شد اما وقت زیادی به دلیل وجود تعاریف متفاوت در دو زبان که باید مشترک می‌شد صرف شد.

5- در برخی موارد باید رفتار کاربر مانند تایپ با کیبورد و موس شبیه سازی شود و نمی‌توان با جاوا اسکریپت کلیک یا متنی را جایگذاری کرد. برای اینکار متدهای ماوس و کیبورد فیزیکی طراحی شد.

6- متدهای کار با NodeList های DOM باید پیاده سازی شود.

7- متدهای کار با فرم، dropdown های انتخابی و دیگر موجودیت های یک صفحه وب باید پیاده سازی شود.

8- مدیریت خطاهای جاوا اسکریپت.

9- در روند پیاده سازی hrobot بر پایه نسخه جدید QtWebEngine به مشکلی برخورد کردیم که ظاهرا در Qt هنوز برطرف نشده، البته در نسخه‌های تست ظاهرا برطرف شده اما تا زمان ارائه رسمی قابل استفاده نمی‌باشد و باید راه حل موقتی برای این موضوع تا زمان ارائه رسمی نسخه بعدی پیدا کرد.

مشکل در مشخص کردن شروع لود یک صفحه یا اسکریپت و پایان آن است، لینک‌ها معمولی مشکلی ندارند اما لینک‌ها جاوا اسکریپتی یا anchor لینک‌ها فقط متد شروع لود را فراخوانی میکنند اما متد پایان لود را فواخوانی نمی‌کنند.

راه حل های ساده زیادی برای رفع این موضوع وجود دارد، اما همیشه و در همه حال قابل اعتماد نیستند، حل صحیح این مشکل به این دلیل اهمیت زیادی دارد، چون پیاده سازی به صورت غیر هم زمان باید انجام شود و اگر جایی با این مشکل برخورد کنیم و راه حل جواب ندهد، برنامه در حالت لودینگ باقی می‌ماند، مگر اینکه صفحه را دوباره لود یا به صفحه دیگری برویم.

از آنجایی که سیگنال loading به کاربر فرستاده نمی‌شود، کاربر باید به صورت polling به صفحه سر بزند و کنترل کند آیا elementای که انتظار دارد load شده است یا خیر. اگر فواصل بین polling 0 در نظر گرفته شود کارایی سیستم پایین می‌آید، اگر بین این فواصل Wait گذاشته شود، از آنجایی که process کاربر و hrobot یکسان است و در دو thread مختلف اجرا می‌شوند wait کردن باعث می‌شود که تمام threadهای فرایند sleep شوند و loading هم متوقف می‌شد، بنابراین هیچوقت این فرایند به پایان نمی‌رسد.

نهایتا برای حل این مشکل thread مربوط به کاربر را درون یک QEventLoop مشغول نگه داشتیم که به صورت polling به thread مربوط به hrobot سر می‌زند. در صورت که load مربوط به element مورد نظر به پایان رسیده بود، از loop تعریف شده خارج می‌شود.

10- برای نصب hrobot روی Fsystem مشکلی وجود داشت زیرا ماژول در thread متفاوتی نسبت به QApplication تعریف شده در process مربوطه اجرا می‌شود و نهایتا Hrobot نیز در یک thread متفاوت نسبت به QApplication اجرا می‌گردد و باعث می‌شود خطای اجرا رخ دهد، زیرا Widgetهای گرافیکی QWebEngine باید در main thread اجرا گردد. برای حل این مساله ماژول‌هایی که از hrobot استفاده می‌کنند به صورت جداگانه در یک sub-process دیگر اجرا می‌شوند و نتایج خود را در آدرسی مشخص ذخیره می‌کنند و SystemApi نتایج را از روی فایل خوانده و به کاربر بازمی‌گرداند.

# **1-5- متدهای موجود در hrobot**

1- متد Save\_as\_pdf

این متد یک export با فرمت pdf از صفحه حال حاضر hrobot می‌گرد.

2- متد Scroll\_to\_end

این متد صفحه حال حاضر hrobot را تا قسمت پایین scroll می‌کند.

3- متد find\_by\_css

المنت‌های موجود در صفحه را از طزیق css selector بر می‌گرداند.‌

4- متد find\_by\_xpath

المنت‌های موجود در صفحه را از طزیق xpath بر می‌گرداند.‌

5- متد wait\_until\_css

تا زمان لود شدن المنت css مورد نظر، کاربر صبر می‌کند.‌

6- متد wait\_until\_xpath

تا زمان لود شدن المنت xpath مورد نظر، کاربر صبر می‌کند.‌

7- متد find\_by\_siblling

Sibling بعدی node ورودی را بر می‌گرداند.

8- متد find\_by\_preceding

preceding بعدی node ورودی را بر می‌گرداند.

9- متد find\_element\_has\_attribute

المنت‌هایی که دارای یک attribute خاص هستند را بر می‌گرداند.

10- متد find\_by\_attribute\_value

المنت‌هایی که دارای attributeی با یک value مشخص هستند.

11- سایر متدها

المنت‌های ذکر شده متدهای کلی هستند و سایر متدها مشتقاتی از این متدها می‌باشند.