Ajith Paul

I am listing two efficient and feasible approaches below. Both the approaches realize the service oriented architecture paradigm.

1. The "Intelligent" Approach

The 3 modules involved are
a) browser extension,
b) data crunching phase
c) dynamic content generation module.
This approach decouples the resource intensive activities to a remote server, thus requiring minimal computing power at the client side (a simple browser).

User experience :

User fires up the browser and he continues his routine web browsing. In the background the browser extension keeps sending the visited URLS to the remote server. As and when the user clicks on the extension icon, he is presented with a dynamically generated web page with the tax implications of his important life changes.

Technical Details :

Browser extension : The browser extension fetches the browsing history of the user and sends it to the remote server (a REST/SOAP call)

At the server: NLP subsystem powered by **Apache OpenNLP** receives the data in the web service.

1. *Domain identification phase* : As a preliminary phase , the site URLs are tested against a predefined regex set. If this phase fails to identify the interest of the user, the system proceeds to the next phase, the data crunching phase.
2. *Data crunching phase* : In this phase, using NLP techniques like **3-medoids clustering algorithm** we map the browsing trends of the user to one of the 3 clusters(home, baby, marriage)
3. *Dynamic content generation phase* : The output the data crunching phase is used by the dynamic content generation module which employs **Python beautifulsoup** to fetch the content from the respective sites (TurboTax blog, TurboTax AnswerXchange and TurboTax YouTube channel) and generate a dynamic page (using **CGI scripts**).

This dynamically generated page is sent from the server and is rendered at the client browser.

2.The Questionnaire Approach

In this approach, the user is presented with a set of questions when the extension is fired and based on the answers provided , the system maps the intent of the user to one of the 3 clusters (home, baby, marriage). Once this is obtained, the system proceeds directly to dynamic content generation phase.