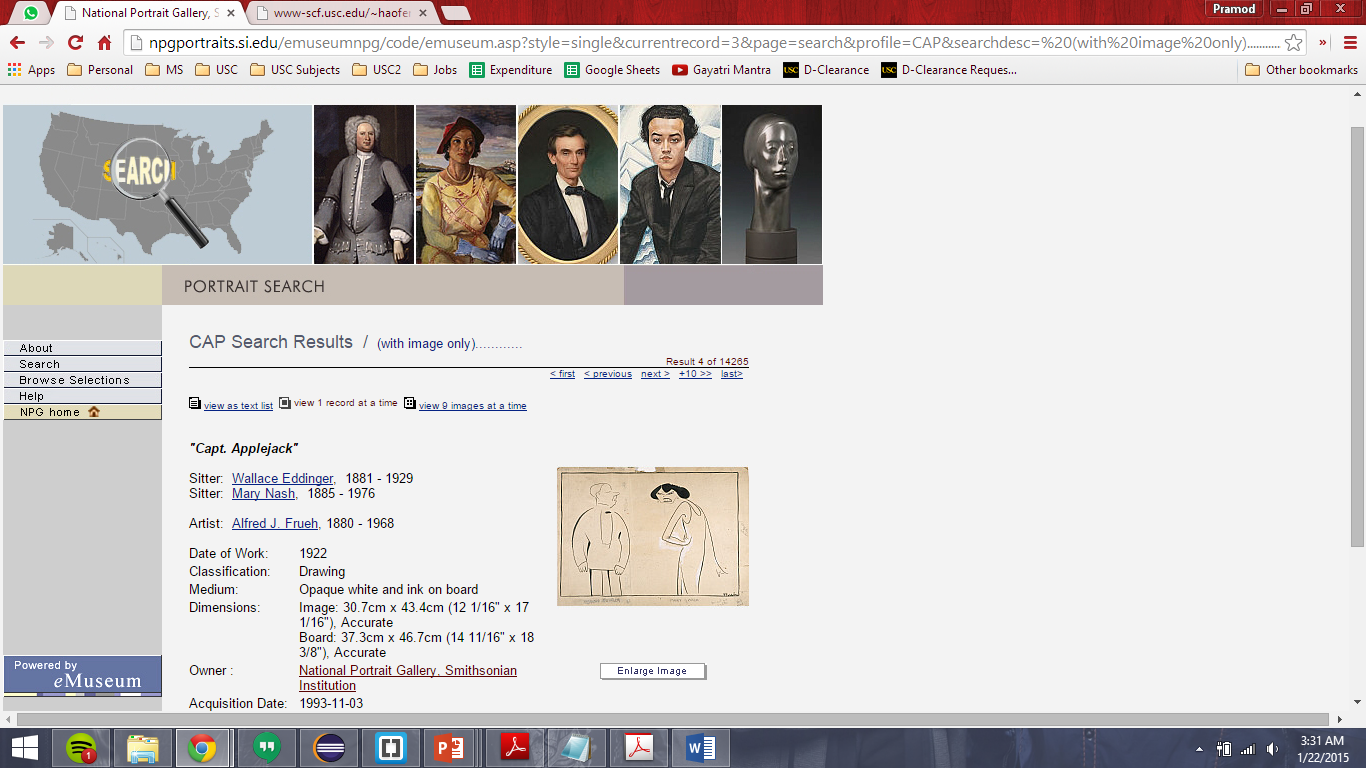
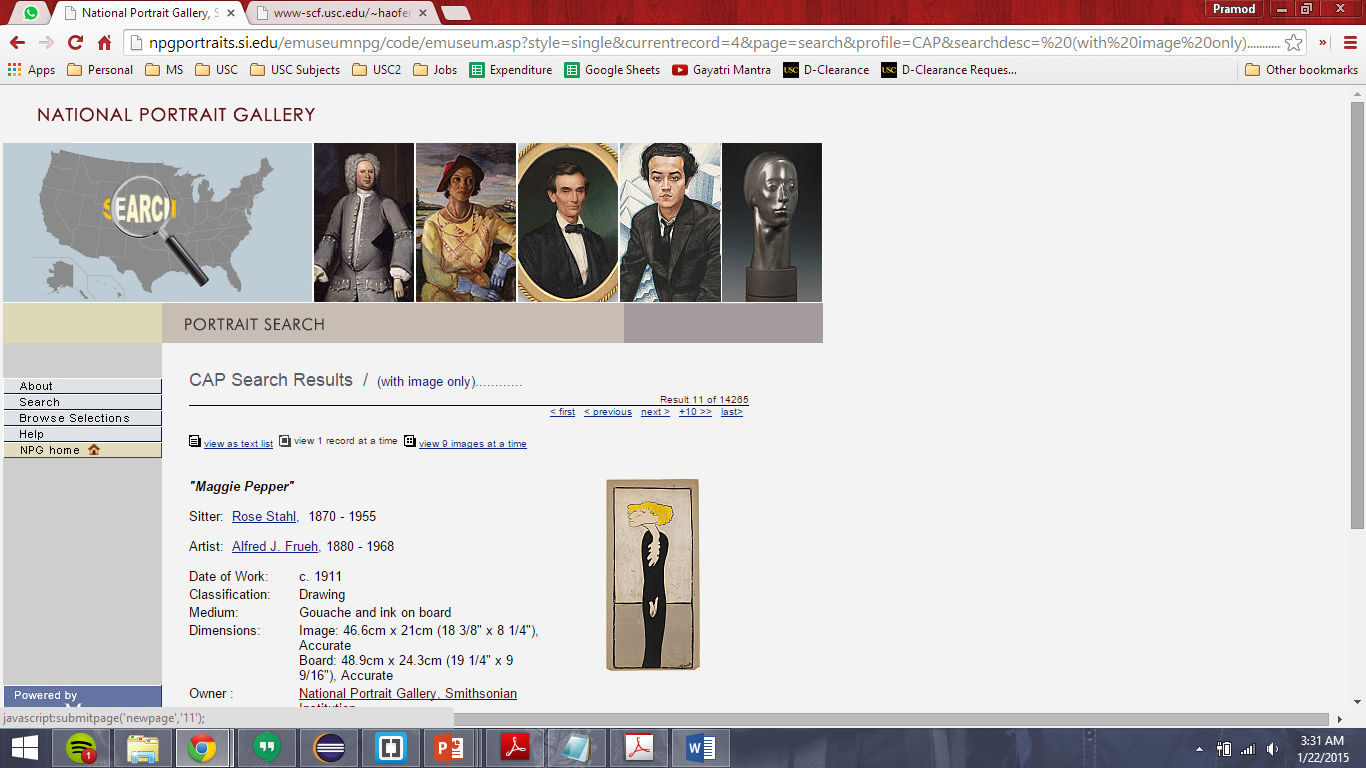
a)

“I Pramod P. Setlur, declare that the submitted work is original and adheres to all University policies and acknowledge the consequences that may result from a violation of those rules”

b)





c)

{

"title": "\"Henry VII\"",

"imageUrl": "http://npgportraits.si.edu/eMuseumNPG/media/previews/9300528B\_1.jpg",

"artist": "Alfred J. Frueh",

"dateOfWork": "1916",

"classification": "Drawing",

"medium": "Gouache and ink on illustration board",

"dimensions": "Image",

"owner": "Unknown",

"aquisitionDate": "1993-11-03",

"creditLine": "National Portrait Gallery, Smithsonian Institution; gift of the children of Al Frueh - Barbara Frueh Bornemann, Robert Frueh and Alfred Frueh, Jr."

},

{

"title": "\"Hunting the Buffaloe\"",

"imageUrl": "http://npgportraits.si.edu/eMuseumNPG/media/previews/9900560B.jpg",

"artist": "Alfred M. Hoffy",

"dateOfWork": "1837",

"classification": "Bound Object",

"medium": "Hand-colored lithograph on paper",

"dimensions": "Sheet",

"owner": "Unknown",

"aquisitionDate": "1999-12-31",

"creditLine": "National Portrait Gallery, Smithsonian Institution; gift of Betty A. and Lloyd G. Schermer"

}

d)

My major technical challenge in this project was with respect to HTMLUnit’s table. The data in the web page I selected was both outside a table and inside a table. The data inside the table could be fetched easily. But the ones inside the table was tricky because of a few hardcodings in the page. All the fields were not consistent across all the paintings. Few were not there and few were. It was crucial for me to use HTMLUnit’s table very carefully to check if the row was really present or not for a particular painting. It had to be omitted if not present but was to be fetched if it was. Other than this, the rest seemed an interesting learning.

e)

I chose HTMLUnit mainly because it had good documentation. I chose to review few tools before choosing this. Tools like JSOUP is relatively new and doesn’t have support for Xpath. HTMLUnit on the other hand does. The intellisense in Eclipse was in fact more than sufficient to use it with ease.

f)

The fields I chose to extract were almost all the fields describing the Image:

Title, ImageUrl, artist, date of work, classification, medium (used to paint the art work), dimension, owner, acquisition date and credit line. I chose to exclude the “sitter” field because many of the paintings didn’t really have this field. And collecting this data and ultimately having many image’s “sitter” as unknown didn’t make sense.