Information Retrieval - Report - Project Art For Sale

Morales Mariciano Jeferson, Piloni Filippo
Universita' della Svizzera italiana, Faculty of Informatics, Lugano, Switzerland

1 Obtained results

In the first period of the project we managed to selected the sites from which collecting the data for the Art-for-Sale project. We choose three main websites:

1. www.artsy.net
2. www.saatchiart.com
3. www.artfinder.com

For each of these sites, we were able to create a spider to crawl the site. We used pre-fixed tags to choose catagories to explore on the sites, accessing to the search page for each of the tag. We then used the crawler to explore the pages of the single artworks and gathering from each the following informations:

- Name of the artwork
- Name of the author
- Url of the image of the artwork
- Price
- Link to the page
- Tags associated to the artwork

- Description of the artwork

We saved the results in json files and we are approaching to start the the creation of an invert index combining all the results obtained. In total, we obtained around 9000 documents to be used in our project

The code can be seen at: https://github.com/JekxDevil/IR-ArtForSale

2 Planned results

The next step for the implementation will be to finish the inverse index and starting the frontend and backend part.

For the inverse index, we decided to index the documents based on the majority of the information, providind a quick access based on tags, price and other characteristics, leaving out from the index only the image, for obvious reasons.

We then decided to implement a simple yet highly customizable interface that will allow the user to filter the results based on the tags associated to the artworks and to get all the important informations about the artworks obtained thanks to the retrieval model. Our plan is to implement at least the following features:

- Result presentation
- Automatic recommendation

With these, it would be possible to accomplish our goal, so to create a retrieve system with a good interface that returns results based on the user's previous researches.